AD-A276 941

(2)

MILITARY OPERATIONS RESEARCH SOCIETY

MORS
MILITARY OPERATIONS RESEARCH SOCIETY

Simulation Validation Workshop Proceedings (SIMVAL II)

Edited By: Dr Adelia E. Ritchie,
Dean, Defense Systems Management College

31 March - 2 April 1992
Institute for Defense Analyses
Alexandria, Virginia

101 South Whiting Street • Suite 202 • Alexandria, VA 22304-3483 • (703) 751-7290 • FAX: (703) 751-8171
e-mail: rwiles@dtic.dgis.dtic.dla.mil

DISCLAIMER

This Military Operations Research Society report summarizes the results of a series of meetings on the subject of verification, validation and accreditation culminating with a workshop at the Institute for Defense Analyses on 31 March—2 April 1992. Each Chapter is authored by the Chair or Co-Chairs of each of the working groups of the workshop and represents the view of that working group and not necessarily the view of the whole workshop. While it is not generally intended to be a comprehensive treatise on the subject, it does reflect the major concerns, insights, thoughts, and directions of authors and discussants at the time of the workshop.

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 1992	Workshop Report 3/31/92-4/2/92

4. TITLE AND SUBTITLE
Simulation Valudation Workshop Proceedings (SIMVAL II)

5. FUNDING NUMBERS
O&MN

6. AUTHOR(S)
Dr. Adelia E. Ritchie, Editor

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Military Operations Research Society, Inc.
101 S. Whiting Street, Suite 202
Alexandria, VA 22304

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Chief of Naval Operations
N81
Washington, DC 20350-2000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unlimited; Approved for Public Release.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This publication represents the proceedings of the Military Operations Research Society (MORS) Simulation Validation Workshop held March 31-April 2, 1992, in Alexandria, VA. This workshop was one of a continuing SIMVAL series that MORS has had in the area of Simulation Validation. It contains the reports of the chairs of the four working groups into which the workshop was organized. It also contains two other reports, one by the overall SIMVAL co-chairs, and one by Dr. Paul Davis which he wrote based on the series activities and his other efforts in verification, validation and accreditation (VV&A).

14. SUBJECT TERMS

15. NUMBER OF PAGES
xiv + 115

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

SECURITY CLASSIFICATION OF THIS PAGE

CLASSIFIED BY:
DECLASSIFY ON:

SECURITY CLASSIFICATION OF THIS PAGE

Military Operations Research Society

MILITARY OPERATIONS RESEARCH SOCIETY

Simulation Validation Workshop Proceedings (SIMVAL II)

Edited By: Dr Adelia E. Ritchie,
Dean, Defense Systems Management College

31 March—2 April 1992
Institute for Defense Analyses
Alexandria, Virginia

101 South Whiting Street • Suite 202 • Alexandria, VA 22304-3483 • (703) 751-7290 • FAX: (703) 751-8171
e-mail: rwiles@dtic.dgis.dtic.dla.mil

The Military Operations Research Society

The purpose of the Military Operations Research Society is to enhance the quality and effectiveness of classified and unclassified military operations research. To accomplish this purpose, the Society provides media for professional exchange and peer criticism among students, theoreticians, practitioners, and users of military operations research. These media consist primarily of the traditional annual MORS symposia (classified), their published proceedings, special mini-symposia, workshops, colloquia and special purpose monographs. The forum provided by these media is directed to display the state of the art, to encourage consistent professional quality, to stimulate communication and interaction between practitioners and users, and to foster the interest and development of students of operations research. In performing its function, the Military Operations Research Society does not make or advocate official policy nor does it attempt to influence the formulation of policy. Matters discussed or statements made during the course of its symposia or printed in its publications represent the positions of the individual participants and authors and not of the Society.

The Military Operations Research Society is operated by a Board of Directors consisting of 30 members, 28 of whom are elected by vote of the Board to serve a term of four years. The persons nominated for this election are normally individuals who have attained recognition and prominence in the field of military operations research and who have demonstrated an active interest in its programs and activities. The remaining two members of the Board of Directors are the Past President who serves by right and the Executive Director who serves as a consequence of his position. A limited number of Advisory Directors are appointed from time to time, usually a 1-year term, to perform some particular function. Since a major portion of the Society's affairs is connected with classified services to military sponsors, the Society does not have a general membership in the sense that other professional societies have them. The members of MORS are the Directors, persons who have attended a MORS meeting within the past three years and Fellows of the Society (FS) who, in recognition of their unique contributions to the Society, are elected by the Board of Directors for life.

PREFACE

This publication represents the proceedings of the Military Operations Research Society (MORS) Simulation Validation Workshop held March 31 - April 2, 1992, in Alexandria, Virginia. This workshop was one of a continuing SIMVAL series that MORS has had in the area of Simulation Validation. It contains the reports (Chapters II-V) of the chairs of the four working groups into which the workshop was organized. It also contains two other reports, one by the overall SIMVAL series co-chairs (Chapter 1), and one by Dr. Paul Davis (Chapter VI) which he wrote based on the series activities and his other efforts in verification, validation and accreditation (VV&A).

Chapter I, Overview, provides the overall SIMVAL approach, its history, the basic definitions and describes an emerging picture of VV&A. *Chapter II, The Basics*, presents a look at three major areas supporting VV&A: documentation, configuration management and independent review. *Chapter III, Verification*, provides an overview of verification and the major methods of verification. *Chapter IV, Validation*, is divided into four parts. *Part I, Validating Models and Simulations*, describes the overall structure for validation, methods and considerations in conducting a validation effort, and a validation documentation approach. *Part II, The Multidimensional Space of Validation*, provides another way of viewing the overall area of validation. *Part III, Face Validation and Face Validity* and *Part IV, Sensitivity Study of a Simulation Model*, describe two validation methods, face validation and sensitivity analysis, and considerations in their use. *Chapter V, Accreditation*, addresses the area of accreditation, its intent, considerations in application and philosophy of use.

There are many different views of the VV&A area. *Chapter VI, A Framework for Verification, Validation and Accreditation* is one of these, which is logical, consistent and held by many in the VV&A community. It is provided to demonstrate that work continues in the VV&A area by many dedicated professionals and by many government and industry organizations.

It is a basic premise of the SIMVAL series that its findings represent a consensus of the military and military support community. Getting a consensus on an accepted VV&A structure will take continued cooperation and support of all those several hundred people who have contributed in the past three years as well as the many others who will participate in the future. This publication is a status report of the SIMVAL series. Future activities will bring further definition and understanding to the overall structure and methodologies. This status report will evolve to represent those new findings.

James J. Sikora
SIMVAL Co-Chair

TABLE OF CONTENTS

APPENDIX

LIST OF FIGURES

LIST OF TABLES

CHAPTER I - OVERVIEW

by Marion L. Williams and James J. Sikora

1.0 INTRODUCTION

Simulation has been an important operations research tool for many years. War gaming has helped develop strategy; campaign models have provided assessments of system utility; engineering simulations have assisted in design of systems and techniques. However, the complexity has changed from simple algorithms to hundreds of thousands of lines of computer code; the number of simulations has grown from hundreds to thousands; and the emphasis has changed from providing insight to that of providing input to decisions on major systems. With these changes, decision makers are asking for some assurance that the models faithfully represent those aspects of the real world that are important to the problem at hand. We have been slower to provide this assurance than we have been to develop new models.

"Validation" is not an easy term to define. It means one thing for the radar range equation or the equations of motion of a satellite. It is quite another thing for interactions in a battlefield. In another dimension, it is less important in some applications than others. For some applications, validation is of little importance; for others, it isn't possible. However, as the results of simulation are presented to support DoD studies, the question frequently asked is: "Has your model been validated?" In the case of operational testing, the question is "Has your model been validated with field test data?" Currently, OSD guidelines on operational testing require that models used to support evaluations be "accredited."

1.1 BACKGROUND

To address these issues, MORS sponsored a series of activities on "Simulation Validation." The series of activities that support the results for this monograph are shown in Figure 1-1. The first activity, a mini-symposium held in Albuquerque, New Mexico, October 15-18, 1990, was hosted by the Air Force Operational Test and Evaluation Center and BDM International, Inc. The mini-symposium provided a forum for general discussion of the broad topic of simulation verification, validation, and accreditation and served as a basis for planning future efforts.

Objectives of the mini-symposium were to:

- Review current efforts in simulation validation;
- Support technical interchange on simulation validation;
- Develop consensus on a consistent set of definitions for terms such as "verification," "validation," "accreditation," etc.
- Develop a plan for future efforts to address issues of simulation validation.

The mini-symposium was divided into five major sessions: Requirements Analysis; System Design; Operational Test and Evaluation; Operations Support and Tactics Development; and Training. Papers for these sessions included case histories, methodologies, lessons learned, and status of current simulation validation efforts.

13 Feb 90	SAG MEETING—PLANNING.
14 JUN 90	58TH SYMPOSIUM—PRELIMINARY COMMUNITY DISCUSSION OF SERIES (WG-28).
15–19 OCT 90	MINI-SYMPOSIUM (SIMVAL I).
16 OCT 90	SAG MEETING 2—REVIEW/DISCUSS DEFINITIONS/ROADMAP.
18 OCT 90	SAG MEETING 3—ESTABLISH DEFINITIONS (V, V & A). UPDATE ROADMAP.
12–13 DEC 90	AD HOC WORKING GROUP I—VALIDATION METHODOLOGY.
7 FEB 91	SAG MEETING 4—REVIEW/DISCUSS ACCREDITATION. UPDATE ROADMAP.
20 MAR 91	SAG MEETING 5—REVIEW/DISCUSS V, V & A.
11–13 JUN 91	59TH SYMPOSIUM—SUMMARY OF FINDINGS/STATUS REPORT.
31 MAR–2 APR 92	WORKSHOP (SIMVAL II)—REVIEW METHODS, DEVELOP BASIS FOR MONOGRAPH.
23–25 JUN 92	60TH SYMPOSIUM—SUMMARY OF FINDINGS/STATUS REPORT (GENERAL SESSION)
3 MAR 93	SAG MEETING 7—REVIEW MONOGRAPH APPROACH/UPDATE ROADMAP.

Figure I-1. Simulation Validation Series Activities

A Senior Advisory Group (SAG), composed of senior analysts representing a breadth of simulation experience, was formed to provide guidance in planning the workshop series, to assist in developing a consistent set of definitions, and to develop a roadmap of activities necessary to arrive at a consensus on a model validation process. The SAG membership is shown in Figure 1-2. The goal of the SAG was to arrive at a consistent set of definitions for simulation verification, validation and accreditation which would be agreeable to all DoD Components, thus resolving the problems caused by the current use of different definitions.

The SAG recommended a subsequent meeting to provide a better description of the validation methodologies. To accomplish this, an *ad hoc* working group meeting was held at The MITRE Corporation on December 12-13, 1990, with DoD Component and industry representatives. The purpose of the meeting was to attempt to define elements of a validation process. Experts in five different types of application areas were invited; Force planning and operations; acquisition; test and evaluation; training; and deployment, mobilization, and sustainability.

The latest session of the SIMVAL Workshop series was held March 31 - April 2, 1992. At this workshop, model verification, validation, and accreditation (VV&A) case studies were discussed, and examples were mapped into the VV&A elements defined at previous meetings. The concept was to use the most pertinent portions of the case studies as examples of specific elements of VV&A.

1.2 VV&A IN THE SCHEME OF PROBLEM SOLUTION

The overall process in which VV&A plays a role is shown in Figure 1-3. The process begins with a problem or set of issues that need to be addressed. Using the scientific method, the problem is decomposed into elements which lend themselves to investigation or analysis. Each of these problem elements can be addressed using different approaches, some of which may include modeling and simulation. For those that are supported by modeling and simulation, a set of requirements for the model to

CO-CHAIRS MARION WILLIAMS, FS, AFOTEC AND JIM SIKORA, BDM

• DAVID ANDERSON, USAF SAA	• DALE PACE, JHU/APL
• PAUL DAVIS, RAND	• NELSON PACHECO, IDA
• HANK DUBIN, USA OPTEC	• JULIAN PALMORE, UNIV. OF ILL
• JIM DUFF, USN OPTEVFOR	• JOHN RIENTE, USA ODCSOPS
• CHRISTINE FOSSETT, US GAO	• DEE RITCHIE, DSMC
• SAUL GASS, UNIV OF MD	• KATHLEEN RUEMMELE, SDIO
• DALE HENDERSON, LANL	• PAT SANDERS, OSD(PA&E)
• RON HOFER, USA STRICOM	• ERNEST SEGLIE, OSD(DOT&E)
• MORT METERSKY, USN NADC	• DENNIS SHEA, CNA
• ALLEN MURASHIGE, USA SAA	• CLAY THOMAS, FS, USAF SAA
• JUM O'BRYON, OSD(T&E/L&MP)	• GENE VISCO, FS, USA MISMA
• MARK ZABEK, IDA	

Figure I-2. Senior Advisory Group Members

correctly and satisfactorily address the element should be developed. These requirements for all elements are grouped together and become the application requirements. The application requirements are then used to compare candidate models' capabilities against in order to select the most appropriate model(s) for the application. The model selection considers not only the model that best satisfies the requirements, but also the credibility of that model for that specific application. This model selection is supported in terms of model capability and credibility by **verification** and **validation**. The results of the model selection process are then used to support an **accreditation** decision. Only after this consideration and a formal accreditation decision has been made, should the model be applied in a simulation, or should model results be used to support an acquisition decision at any level.

1.3 DEFINITIONS

Basic VV&A definitions were developed during initial SIMVAL meetings, and then used throughout the workshop series. There was nothing dramatically new in the definitions; they were modifications of those currently being used by some organizations. However, they were fully discussed and honed until a consensus was reached. Other definitions could have been chosen which are adequate. However, the goal of the SIMVAL series was to agree on a common set of definitions so that we could more clearly and easily communicate.

The following set of definitions was developed by the SAG and agreed upon by the SIMVAL participants:

> VERIFICATION: The process of determining that a model implementation accurately represents the developer's conceptual description and specifications.

Verification consists of two basic types. Logic verification ensures that the basic equations, algorithms, etc., are correct. Code/object verification ensures that these representations have been correctly implemented in the computer code.

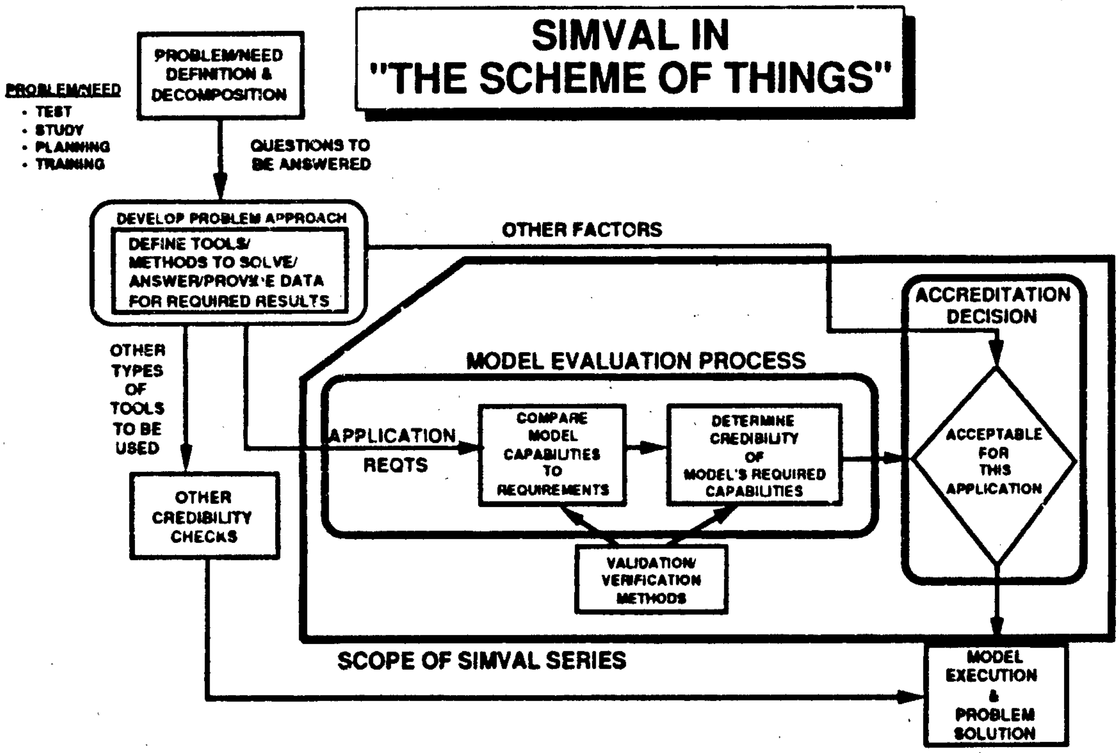

FIGURE I-3. VV&A in the Problem Solution Process

> VALIDATION: The process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model.

The primary change in this definition is to recognize that validation is not an event, but a process consisting of several steps to measure the degree to which the model represents the real world. These steps may consist of face validation, data base validation, etc. Complete validation, i.e., ensuring that the model represents the real world in all aspects, can be achieved only for simple models, since complete validation implies that the model can be used for *any* application. A complex digital simulation can achieve degrees of validation, but complete validation is a goal that can probably never be reached. Therefore, it would be improper to refer to such a model as "validated."

> ACCREDITATION: An official determination that a model is acceptable for a specific purpose.

Accreditation is a decision that is based on a number of different factors, including V&V. It accepts that a given level of V&V is sufficient for a model to be

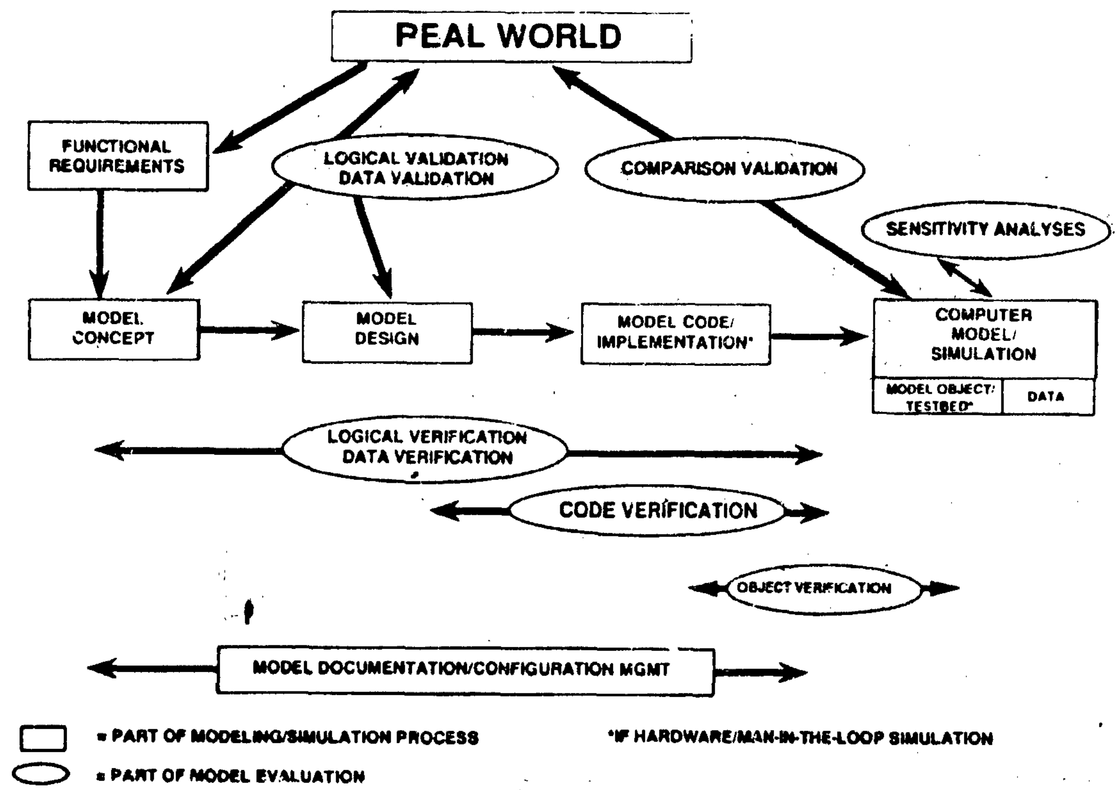

FIGURE I-4. The Relationships of V& V to the Model Form

used in a particular application. For some applications, a low level of V&V (for example, code verification) may be acceptable. For other applications, a more rigorous validation may be necessary. Accreditation must take into account the importance of the decision in determining the rigorousness of V&V required, as well as other factors.

1.4 THE VERIFICATION AND VALIDATION STRUCTURE

Verification and validation are comparison processes. Verification compares the implementation of a model against the intent or design of the model. Validation compares the model against the real world. Both are processes that establish the credibility of the model in performing certain functions. Accreditation, on the other hand, uses the credibility of the model in a formal decision as to whether the model can be used for a specific application. The relationships of verification and validation are shown in Figure I-4.

The "Real World" at the top of the figure denotes the actual function or system which is being modeled. If part or all of the function or system does not exist (e.g., a future aircraft), then it is our best realization or understanding of that non-existent function or system. From the Real World, the model designed selects the functions and systems that are important to the class of problems that are the intended application set for the model. This then is the basis for the **Functional Requirements.**

These requirements reflect the types of functions or systems to be modeled (e.g., ECM, command and control, M1A2) as well as an indication of the level of detail desired (e.g., signal level, message level, operator level). The model requirements become the first category of model structure which falls under the need for **Documentation** and **Configuration Management**. For a further explanation and description of documentation, refer to Chapter 2. Model requirements should be documented when developing the **Model Concept**.

The **Model Concept** is an initial model architecture which gives a consistent and sufficient relationship description between the functions and systems to be modeled. The model concept should satisfy all the functional requirements the modeler defined earlier. The model concept is then documented (and thereby falls under configuration management) and becomes the basis for the **Model Design**.

The **Model Design** is the structural outline of the model. It defines the modeled system elements, their functioning, and their interrelationships. This design can be done in successive levels of detail until sufficient definition is available to translate into the **Model Code**.

The **Model Code** is the model in its computer language form. The Model Code finally is compiled or assembled into its computer instruction form, the **Model Object**. This, along with the data required to execute the model and to simulate the desired situation, is the **Computer Model**. The Computer Model (or some functional elements of the Computer Model) may be implemented in hardware or performed by humans as part of the simulation.

The verification processes then are the checks made at one stage of model development against the requirements, design, and/or form of earlier stages to assure correct translation. For example, code verification is the process of comparing the model code stage against the requirements and specifications of the model design to ensure the code correctly represents the design. For further explanation and description of verification processes, refer to Chapter 3.

The validation processes compare any stage of model form against the real world. For example, comparing the model design against the functionality of the real world system can help ensure that the design represents all the necessary functions of the real world to satisfy the uses of the model. Another validation method would be to compare the output of the computer model against the functional performance of the real world system under the same initial conditions. For further explanation and description of validation processes, refer to Chapter 4.

The credibility that a model gains by applying verification and validation processes is part of the input to the accreditation decision. This was shown earlier in Figure 1-3. Accreditation is addressed in more detail in Chapter 5.

CHAPTER II - THE BASICS

by Joseph J. Cynamon

2.0 INTRODUCTION

The basics of model or simulation verification, validation, and accreditation (VV&A) are the methods and tools used to track, record, and control the model development and VV&A processes. This chapter discusses some of these methods and tools: documentation, configuration management, and independent review of the model.

The message delivered by this chapter is that VV&A is composed of a series of tasks that contribute to its accomplishment. If any of these tasks or elements is not successfully accomplished and completed, then the VV&A enterprise can also be expected to fall short of its goals. VV&A can be thought of as a sum of many elements. If any of the links in this chain of elements is violated, VV&A will be endangered.

This chapter covers in detail the most important elements of these basics. Figure II-1 is a summary flow chart used to identify some of the basic elements crucial to VV&A. It also highlights the feature that each element contributes.

- **Documentation** provides the description of the model or simulation, its requirements, how it operates, and its characteristics, algorithms, and intended application(s). Documentation should describe the history of the development of the model and the methods used for testing its functionality and properties.

- **Configuration Management** (CM) provides for the tracking (i.e., an audit trail) of the development of the model or simulation. Each of these functions provides an

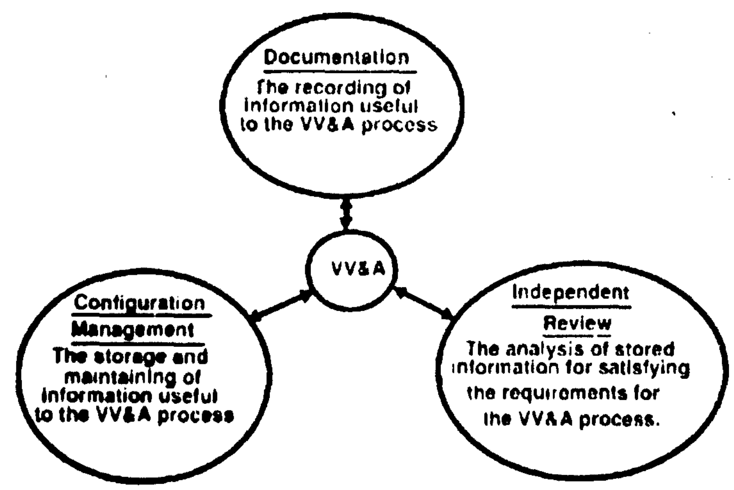

FIGURE II-1. The Basics Chart

information source for VV&A.

- Finally, **Independent Review** is the process in which an impartial expert reviewer(s) conducts a critical evaluation of both the product and the VV&A process performed on it. This review should be done without bias and reservation, and must be conducted independently of the influence of the product developer(s). The independent reviewer(s) should have full access to all documentation and the cooperation of the developer(s) and VV&A participants. This means availability of all levels of documentation, total availability of the configuration resources, and complete cooperation of the participants for consultation with the reviewer(s).

2.1 DOCUMENTATION

The definition of documentation developed by the SIMVAL working group follows:

> DOCUMENTATION: Analyst's manual, user's guide, programmer's manual, etc., providing the math, program structure, assumptions and algorithms used, including documentation of procedures and results of any verification and validation efforts.

The function of documentation, under VV&A, is to provide information about all aspects of a model's intended application(s), description, and history to the professional community.

The flow chart in Figure II-2 identifies the critical steps in the documentation process. In this diagram two feedback loops are identified. The first loop illustrates the process of establishing documentation requirements based upon the development of the evolving VV&A requirements. The second loop is also a feedback process based upon the program's progress and the implementation of VV&A. Its purpose is to maintain responsiveness to the review processes as additional documentation needs develop.

Given the above definition of documentation, Table II-1 details some of its critical elements and identifies the product or process needed to implement them.

2.1.1 Documentation Application Techniques

Factors that have the largest impact on VV&A documentation support include the following:

- Traceability to model requirements.
- Description of the functional design of the model.
- The identification of the models and algorithms used in the specific application.
- Data requirements identified for VV&A.
- Confirmation that the required aspects of VV&A are fully covered.
- A record of the history of the product's design, development, testing and VV&A.

Of primary importance is traceability. Historical records should be maintained to trace requirements with product development (e.g., to link test procedures to requirements to demonstrate they are met). Documentation provides the means to record and review the goals and objectives of model development, theraceability of test

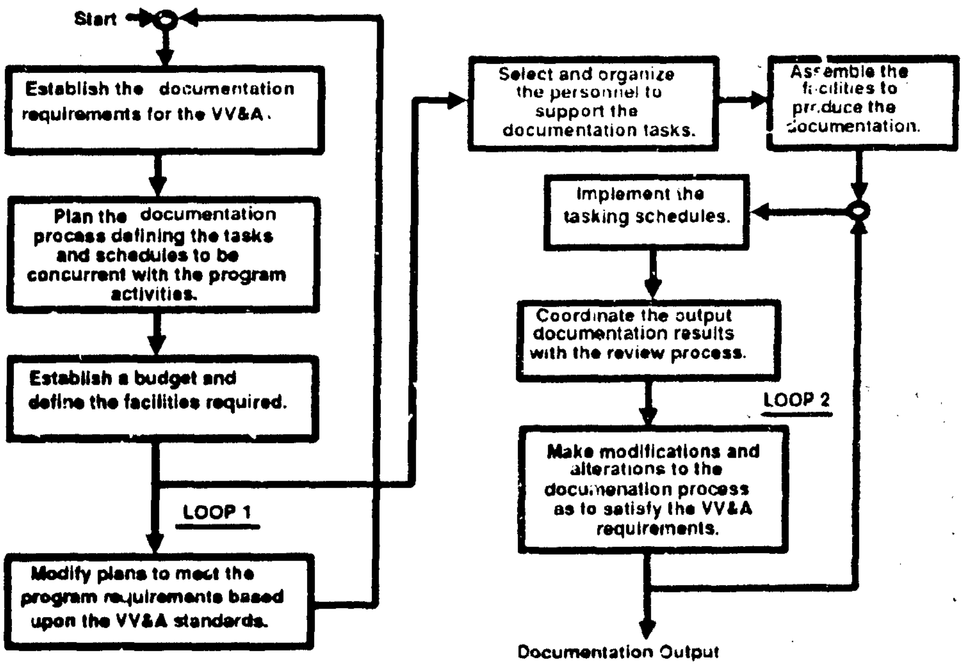

FIGURE II-2. Critical Steps in the Process of Documentation

results with performance goals, and the fulfillment of functional requirements.

A fundamental requirement is that an adequate detailed description of the model be given in its documentation. Is there enough of a description to meet the requirements disclosure? Are changes made to algorithms and math models recorded completely and in a timely manner? Have coding revisions and changes been reflected with proper comments and logged in the source code and accompanying documents? Are revisions formally noted and controlled in the documentation so that users are properly alerted? Have the developers published their findings of verification testing of key parameters so that they can be traced to the proper performance requirements?

Table II-2 provides a list of minimum requirements for verifying traceability of VV&A documentation. This list can be used to check for documentation items that can be regarded as a minimum functional package to support VV&A.

Complete and understandable documentation of the model's functional operation is essential. Designing documentation that describes functionality takes creativity and initiative, keeping in mind that information must be sent to a user or to the application community in concise and understandable language. It is recommended that liberal use be made of illustrations, diagrams, and charts. It is also recommended that functional block diagrams organized into multiple levels for systematic detailing

Table II-1. Documentation Needs

ITEM	DESCRIPTION
Technical description	Manuals, text books, diagrams and charts, pictures and illustrations, and photographs.
Description of uses	Demonstrations, training courses, user and applications examples, on-line user manuals.
Installation procedures	Installation procedures, automated installations processes, vendor-supplied support.
Installation validation	Examples to exercise critical functions and I/O supplied data functions for automated installation comparisons.
Model design requirements	Design goals and specifications that are criteria for VV&A evaluation and acceptance.
Algorithm and math Model	Manual containing the algorithm reference sources, assumptions and known limitations, and any supporting performance analysis.
Computer program design	Description in words, diagrams, charts, tables, and pictures both in hardcopy and in electronic storage.
Testing of the model	The purpose and description of the tests, the key parameters to test, the analysis for setting the acceptance of performance, the ranges of the acceptance of the results and the testing procedure description.
Training of operators	Instructor-lead course materials, computer programmed instructive course, manuals and charts.
Dictionary of terms	Definitions of terms and parameters, tables of interrelationships of variables, arrays, and computer subroutines.

of the model be used. At the top level, the diagram can begin with a general architectural description of the model. This can be followed by a detailed tree diagram describing the sequence of executable routines. A functional block diagram can then follow, describing the sequence of operations performed through the various modes and processes during execution. Other useful diagrams are looping charts, which identify the logical functions controlling the branching of sequence of the model processes and condi-

Table II-2. Recommended Minimum Requirements For Traceability

•	A checklist of design goals and objectives for the model.
•	A dictionary of key parameters.
•	A matrix that identifies quantitative performance values and their acceptance ranges correlated with the requirements.
•	A detailed description of all algorithms and math models used.
•	A list of all tests performed and their results in a matrix that correlates these results with the requirements.
•	A functional description of the model that can be correlated with the requirements and specifications.
•	An historical listing of design and development modifications made to the source code, and the rationale for the changes.

tions for switching. Looping diagrams can show parameter and variable processing through the subroutines of the model. They identify the interface details and many of the key transfer characteristics processes. In addition, they can be made to provide many otner key computational characteristics and functional routines, such as array size, word length, processing speeds, and processing conditions. Signal flow diagrams also may be used to describe operations quickly. They can be used to quickly point up pertinent model characteristics to the reviewer(s). The process can easily be supported by text descriptions that provide additional details about each element in the diagram. Table II-3 summarizes the aforementioned diagrams, describing their functions and some key characteristics they provide.

Spread sheets provide another useful format for organizing model descriptions. Subroutine calls can be itemized with a brief functional description. These types of charts could be used to identify each routine, the routine that calls it, and the routines that it calls. It could introduce the arguments passed to and from the routine, and identify the parameters passed to and from it through common blocks and structure statements. The advantage of this format is that various sorting techniques can be automated to recover the specific interrelationships of structure and characteristics.

Additional computer-automated tools also can be used to analyze and produce documents that describe extended and complex models. Some of these tools produce tables that catalog routines and their interfaces, their parameters, their arrays, and their variables associated with the appropriate subroutines. They can analyze the program's operation and timing, searching for programming errors and warning of possible conflicts. They also analyze the program's statistics and processing efficiency, making recommendations for improving the running performance. The information collected can be stored in separate files for later review, in support of VV&A, and can be used to sort through the tables for specific identification of the program's interrelationships.

Table II-3 Summary Table of Block Diagrams

ITEM		FUNCTION	CHARACTERISTICS
1.	Overall architecture	Provide a list of processes used, the order of their use, and the conditions for their selection.	Show: preprocessor for conditioning input data, command file for executing the primary program, post processor for interactively conditioning the output file, graphic program for producing user information and observations, and interactive integrator for allowing users to interface with program's execution.
2.	Tree	Identify subroutines used in all of the program phases, showing sequencing and conditions for selection.	Program phases are: Initialization, input reads, opening of the output data files, ordering the processes and rules of performance, the operational phase, the termination and file closing phase.
3.	Functional block	Describe operation sequence, modes, and processes performed.	Radar simulation would have a sequence of operations that include the waveform generation, the transmitter power modulator, the antenna scan and coverage processing, the receiver, the IF mixer, the signal processing, etc. The system modes might include the search phase, an acquisition phase, and a tracking phase.
4.	Looping	Highlights logical functions and branching conditions and sequencing based upon inputs.	These are specialized diagrams that might provide useful information to correlate with the system requirements.
5.	Signal flow	Provide observers with direct contact with the code operations that include insights to functions, logic, and branching of signals being processed.	This diagram might provide the operation of routines, identifying their functions, the parameters they are processing and their parameter characteristics.

Table II-4. Suggested Format for Historical Records of Model Testing

ITEM	DESCRIPTION
Purpose of tests	Identify the test requirements from the design specification.
Description of test	Describe the method(s) to be used in performing the test, including: the dynamics, observation intervals, range of values to be tested, the parameter constraints of the input to be applied, and how they were controlled.
Record the results	Record results as they occur, over time and space, recording all conditions and constraints observed during testing.
Post the Analysis	Provide all analysis computations performed, including the algorithms and the math processes used. Present results in the format as specified by the documented requirements.

Comprehensive historical documentation for all tests performed during the product's development, including a description of the purpose of the tests, the test results, and the analysis of these results, is essential. Historical documentation will tie testing to the specific design requirements of the product. All of this should be incorporated into the design and development logging document that will be one of the bases for verification and independent review. The document must be archived for future review by developers and users as a basis for additional new model applications.

Table II-4 suggests a format for the historical recording of the development testing requirements. It is not complete but does identify the functions and descriptions of some major elements.

The documentation should describe the math models and algorithms used. In describing these, connections should be made to the coding process used, to the assumptions and conditions for them to be appropriate, and to the operational modes for which they are applied. For instance, to calculate a state vector of any object, a numerical integration process is required. The implementation of that process should be discussed and its selection justified. If coordinate frame changes are needed, a discussion of the choice of frames should be given with the derivations of their equations. In addition, the use of classical performance relationships should be specified. For example, in the use of the radar range equation for predicting radar performance, the parameters and their appropriateness to that application should be justified. It is recommend that an analysis manual be organized around the use of functional flow diagrams.

Table II-5 highlights the types of descriptive information that a reviewer needs to assess design adequacy.

It is also recommended that a set of documents be created specifically to support VV&A. This set of documents includes the verification and validation plan, verification and validation report(s), and any accreditation decision and decision support documents for previous applications of the model or simulation. Further details of the con-

Table II-5. Checklist of Model Description Entries

•	Develop a systematic list of model operations performed and identify the algorithms and math models used.
•	Each of the algorithms and math models should have identified references justifying their use, or derivations from known and sound physical principles. All assumptions and limitations must be identified along with any analysis of applicability.
•	To support algorithm analysis, diagrams identifying time and space parameters and their functions must be included.
•	The coding implementation of these algorithms should also be described. An analysis investigating the computational accuracy supporting the coding algorithm should also be given.
•	Conditional descriptions for the computation must also be included. This includes coordinate computation frames, time lines, conditional events and modes, and sequences that are conditioned on events.

tents of these documents are discussed in Chapters 4 and 5.

2.1.2 Documentation Strengths

VV&A may be improved by imposing and enforcing sound documentation principles. Some of the advantages gained by maintaining documentation are identified as follows:

- Having documented records of the products developed and tested ensure traceability to requirements.
- Providing resources for storing and maintaining library facilities in a designated archive with defined control procedures assures information security.
- Setting minimal standards for describing and updating model applications assures that historical tracking and extended life potential will accompany model reuse.
- Establishing a facility to store documented records of VV&A events and conclusions provides future applications with traceability of VV&A.

The location for storing documentation should be central and controlled to assure protection against loss and tampering, and accessible to all model reviewers and users. Once a document is approved for archiving, it becomes protected and only the designated control group is authorized to make and promulgate revisions.

Table II-6 summarizes the essential elements of archiving in terms of the reporter's who, what, when, where, and how.

2.1.3 Documentation Limitations

The limitations that can have an impact on documentation include:

Table II-6. Essential Elements of Systematic Archiving

QUESTION	ELEMENTS
WHO?	
a.	The model procurer
b.	The developer
c.	The user
d.	The independent reviewer
e.	The accreditation team.
WHAT?	
a.	Requirements, plans, budgets, identifications, schedules, and applications.
b.	Description, operation, design, test procedure, test data, and reports/reviews.
c.	Applications, user form reports, improvements, and operational limitations.
d.	Information used in reviews, reviewer notes, conclusions, and recommendations.
WHEN?	
a.	Start at the program outset through all phases of program development and applications.
b.	Design phase, integration, testing, and engineering development.
c.	User group meetings, application experience during testing, and user improvements and extensions.
d.	During planning and scheduling, data collecting, and reporting.
WHERE?	All elements become archived by configuration management team in their facilities.
HOW?	All elements should be stored in both electronic and hardcopy forms.

- Model or simulation development programs do not always provide sufficient funding, schedule, and manpower for documentation.
- The caliber of documentation often depends upon the caliber of the program requirements.
- Inadequate documentation makes VV&A more difficult.
- Poor documentation is almost always accompanied by poor configuration management, and vice versa.
- When program cutbacks occur, documentation and configuration management are usually the first to be reduced.

Validation and accreditation rely heavily on the quality of the documentation of the model. VV&A must be supported with real project dollars, near-term and long-term scheduling to integrate the documentation with each phase of the program, and dedicated resources and staff.

One continuing difficulty in setting up documentation requirements for hardware procurement programs is that many engineers do not take the time to document their work. However, it is expected that new emphasis will be placed upon making time for documenting one's work, and the technical staffs supporting the model will require good documentation skills.

2.1.4 Documentation Lessons Learned

From looking at projects that did receive complete documentation as well as those that did not, the following are some of the lessons learned at the SIMVAL workshop series:

- If the model is not documented with comprehensive top level overview descriptions, the review process is more difficult.

- The lack of historical records of the product's development defeats the traceability required by VV&A.

- The model's documentation must be tailored to the application's requirements and must describe the specifics of assumptions made.

- The documentation must incorporate inputs and suggestions from both the developer and the users.

- The support documentation also must focus on user needs. These include describing the installation and operation of the product, descriptions of potential machine hardware impacts on hosting the product, and the supply of a typical example for turnkey operation checks.

- In the practice of planning a product's development, funding for the documentation must begin at the start of the program.

Many have experienced model documentation that provided volumes of detail but lacked a useful top level overview or executive summary. The overview needed should answer the following questions: What does the model do? How will it function? What does it need as inputs? What will it provide as outputs? What are its internal functions doing? With what level of fidelity does it perform these functions? How have these functions been shown to be representative of the actual application? What is needed to use this model? What assumptions were made? What factors that cause deviation were neglected?

2.2 Configuration Management

CONFIGURATION MANAGEMENT (CM) is a discipline applying technical and administrative oversight and control to identify and document the functional requirements and capabilities of a model and its supporting databases, control changes to those capabilities, and document and report the changes as required by VV&A. Configuration management includes ensuring the detailed design and the computer source code of the model are properly documented and tracked.

The most important responsibility of the CM group will be the maintenance of version control of source code for the project.

In Figure II-3, a flow chart is presented that highlights some of the principle processes over which configuration management has cognizance, including:

- The storage and maintenance of the model requirements,
- The storage and oversight of the model revisions and upgrades,
- The library/archiving of descriptive materials, records, reports, and program documents.

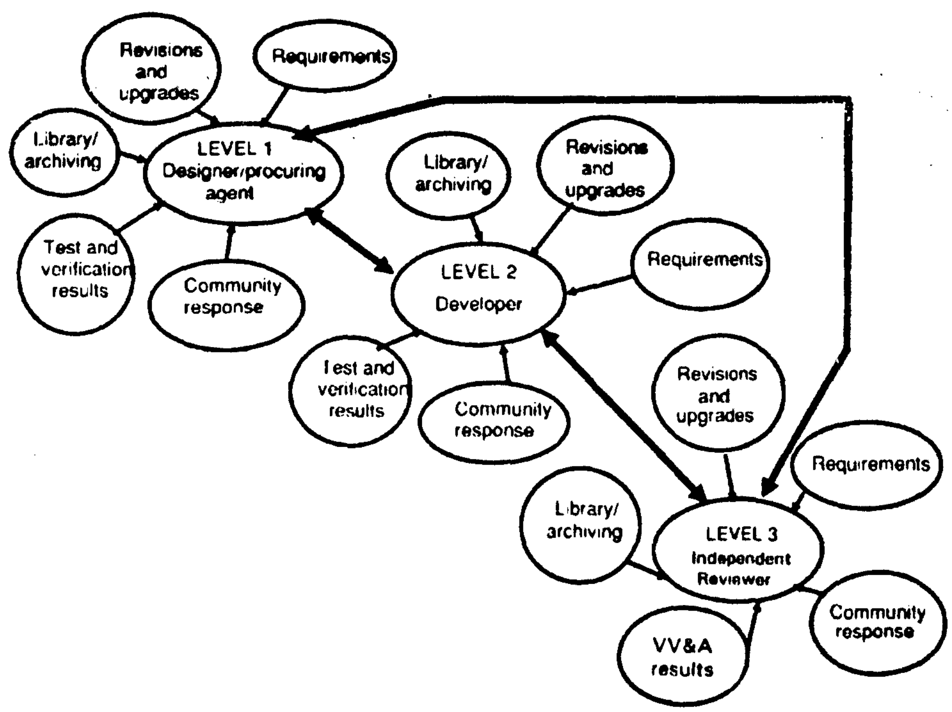

FIGURE II-3. Levels of Configuration Management

The process of configuration management can be performed at least at three interest levels: the designer/procurement level, the developer level, and the user community level. Although all levels are shown to share common interest in the five basic functions, the perspective and actual content of materials emphasized for each level may differ slightly. Although the users of the CM facility are shown as separate entities, all require access to subsets of the same set of documentation maintained by the configuration manager.

Implementing a CM group's policy requires that project support and funding be established at the beginning of the program. Rigorous procedures must be adopted and strictly adhered to. Administration of these policies should be exclusive of other working program functions in order to exercise independence. Items to be managed and archived by the CM group should include: model requirements and specifications; model source code, description and documentation; test plans and reports; operating procedures; input and output database information; and performance information.

Tracking and archiving the historical development of the product is another important function served by the CM group. The most important responsibility of the CM group will be the maintenance of version control of source code for the project.

Table II-7 presents a summary of the elements and a functional description of CM in support of VV&A.

Table II-7. Essential Elements of Configuration Management for VV&A

OBJECT	DESCRIPTION
Requirements	Requirements, specifications, model description, and goals and objectives of the model design.
Descriptions	All documentation that describes the functions, operations, and the testing to verify the model design meets requirements.
Model Revisions	Maintain independent archiving of code and upgraded versions to track program progress.
Test and Verification	Maintain archive of all testing procedures and results from all phases of the model development, modification, and application life cycle.
Library Correspondence	Maintain an archive that traces the history of the model and model reports from design and development phases, through user applications, reviews, and community experience.
VV&A	Archive all plans, reports, correspondence, reviews, and findings associated with VV&A.

2.2.1 Configuration Management Application Techniques

Items that are critical to maintaining configuration management include:

- Maintaining archives consisting of I/O databases and test data collected during developmental and verification phases.
- Tracking and archiving the results of the overall VV&A, recording each project phase and VV&A event.
- Archiving all project requirements and measures of performance.

A clear definition of the configuration management process for the project is required as part of the archival collection. It should express the ranges of responsibilities assigned the CM group and the functions that the group must perform to support the project. The CM group will normally create its own archival procedures, computer architecture, and file structures. It will create a document defining the methods it adopts so all project activities can review and understand the CM environment and its information requirements.

Exceptions to the data format control by the CM are the database storage and test data formats. These formats are created by the developer, user, and the VV&A communities t. ʼit the needs for their responsibilities. The CM will be responsible for accepting this data and creating an archival environment to maintain and protect the contents of these files without alteration, for future retrieval and assessment. The I/O and test data archives should be available to all project members, subject to security sensitive need-to-know clearance.

The CM will maintain the results of VV&A by tracking and archiving each VV&A milestone or event. This information contains V&V plans and reports, accreditation decisions, and the results of all independent reviews.

2.2.2 Configuration Management Strengths

The CM improves VV&A in the following ways:

- A complete historical record of the model's development and application history facilitates VV&A.
- The CM's automated facilities are responsive and flexible in responding to inquiries.
- The CM provides documentation security.
- The CM can maintain support for multi-level secure versions.

The advantages of using centralized computer storage by the CM to track the model programs are clear. The process of keeping a historical log of the various versions of the project's model, independently of the developers influence, creates opportunity to protect against file losses and corruption. The automated storage environment of documentation makes timely access to information within reach of the other project members.

Another strength associated with an independent CM group is the ability to provide documentation security, avoiding tampering with or loss of information needed in VV&A. A systematic control of the filing of information, isolated from the project's routine activities, protects against data loss and contamination.

The tracking of test results with the model version used to create the data is a significant strength of the CM facilities' support of the review process. The storage of older versions in a secure environment is the safety net for catastrophes that could and do befall the projects. Also, having archived multiple versions of software can allow users and reviewers to reconstruct the rationale for past decisions.

2.2.3 Configuration Management Limitations

Some of CM's limitations include:

- The CM group has only the Government software code requirements and standards to follow, instead of broader system standards.

- The CM needs a repository for the model and the documentation, separate from other project activities.

- The CM group has not always been recognized as an essential activity in the support of VV&A.

The CM group must follow existing software specifications that are defined by the government regulations. However, these software specifications alone are insufficient to accommodate the information needs demanded by VV&A.

The cost of configuration management is increased by of the need for separate storage and operating facilities. The argument must be made that project isolation and security justify the investment. If VV&A is to be established as a qualification of a model's application, then a CM facility must be accepted.

2.2.4 Configuration Management Lessons Learned

The following are lessons learned about configuration management and VV&A:

- The CM process must be allowed to evolve over the program's development and must review materials for backwards compatibility (i.e., must "benchmark" materials).

- The CM must maintain materials in such a form that they can be traceable to program design requirements.

- CM facilities should be maintained to support the three interest levels of program activities and should be tailored to support functions applicable to each of these three interest groups (see Figure 2.3).

- The CM should be endowed with facilities to support his assigned responsibilities and have the authority to carry out these responsibilities.

- The CM should maintain close contact with the user community by participating with user groups and developing a repository of user lessons learned.

- The CM should maintain a knowledgeable staff that can support adequately the potentially immense library of technical materials supporting a model or simulation.

The CM must be sensitive to the application process of the model and recognize the user interaction. Once the product is released to the community, the scrutiny it receives will require much more configuration control. The CM must provide expanded services to the users, understand their viewpoint and application evolvements, and support the cataloging of their experiences. The archiving of the user's data and documentation (reporting their experience and tracking any changes made in revising the product) will be a significant function of the CM. It will require the CM to participate in user workshops and working groups, and gain cooperative support from the user. The CM should maintain files of the user versions, and expect to create documentation to support these revisions. When possible, the developer should be keep abreast of developments by the CM, and vice versa.

In our experience, unless the program has funding to maintain developer involvement for upgrading, the developer usually steps out of the program. This leaves the program office to manage the configuration. Unless the project organization recognizes the need for formal configuration management, it may not happen. Because many program offices have continuous personnel changes, the configuration continuity is soon lost. With the requirement for VV&A, the configuration management process will be required throughout the life of the product and must be planned well and funded adequately.

2.3 Independent Review

INDEPENDENT REVIEW is performed by competent objective reviewers who are independent of the model developer. It includes either (a) a detailed verification and/or validation of the model; or (b) an examination of the verification and/or validation performed by the model developer.

Figure II-4 summarizes the information an independent review team needs in order to make a reasonable assessment of the model. Together, these items compose the processes and functions that have been discussed in the previous sections on documentation and configuration management. They are seen as essential information inputs to the independent reviewer, and if an attempt to accredit a model is made without them, the process would be much more difficult.

Additional discussion on the review processes that lead to accreditation is contained in Chapter 3, "Verification," Chapter 4, "Validation," and Chapter 5, "Accreditation."

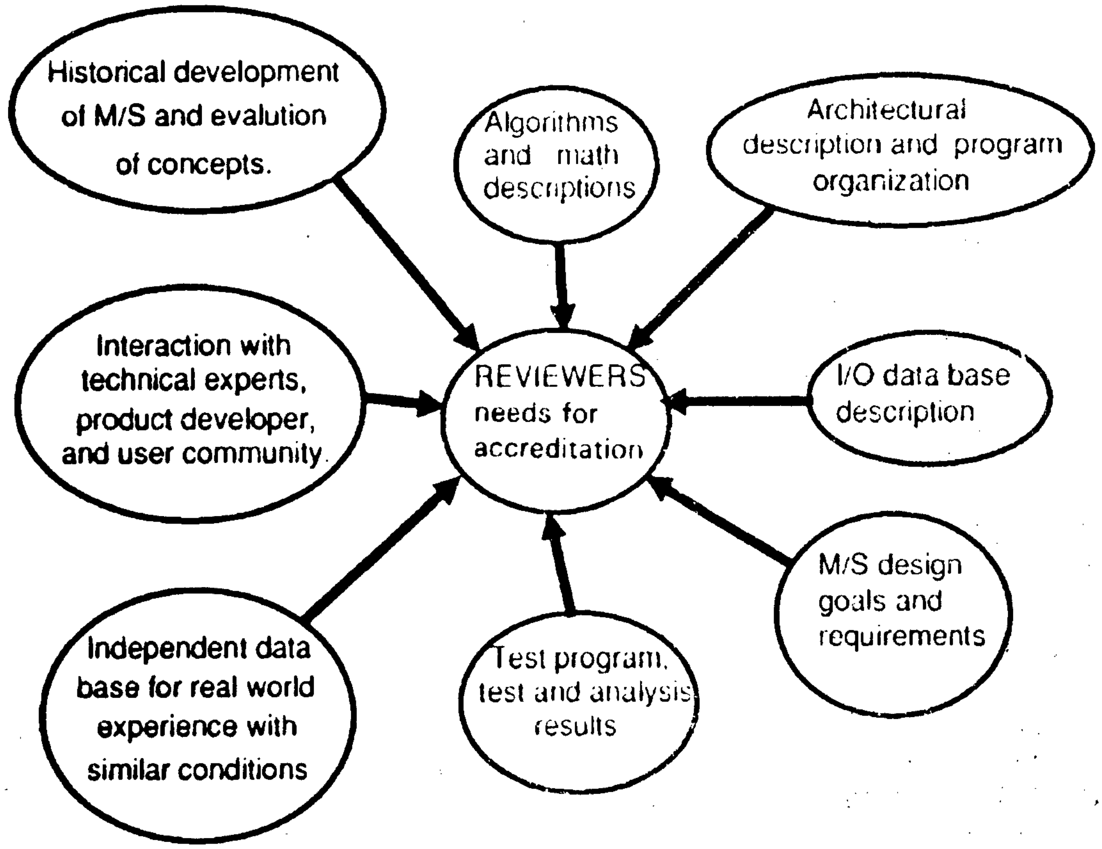

FIGURE II-4. Information Resources Needed by the Independent Review Team

CHAPTER III - VERIFICATION

by Jim Metzger

3.0 INTRODUCTION

3.0.1 Overview

> **VERIFICATION is the process of determining that the implementation of the model or simulation accurately represents the developer's description and specifications.**

Figure I-4, in Chapter I, illustrates how verification fits into an overall V&V framework. This figure is reproduced below for the reader's convenience. The primary verification methods — logical verification, code verification, data verification, and specific logic or assumption comparison — are discussed in this chapter.

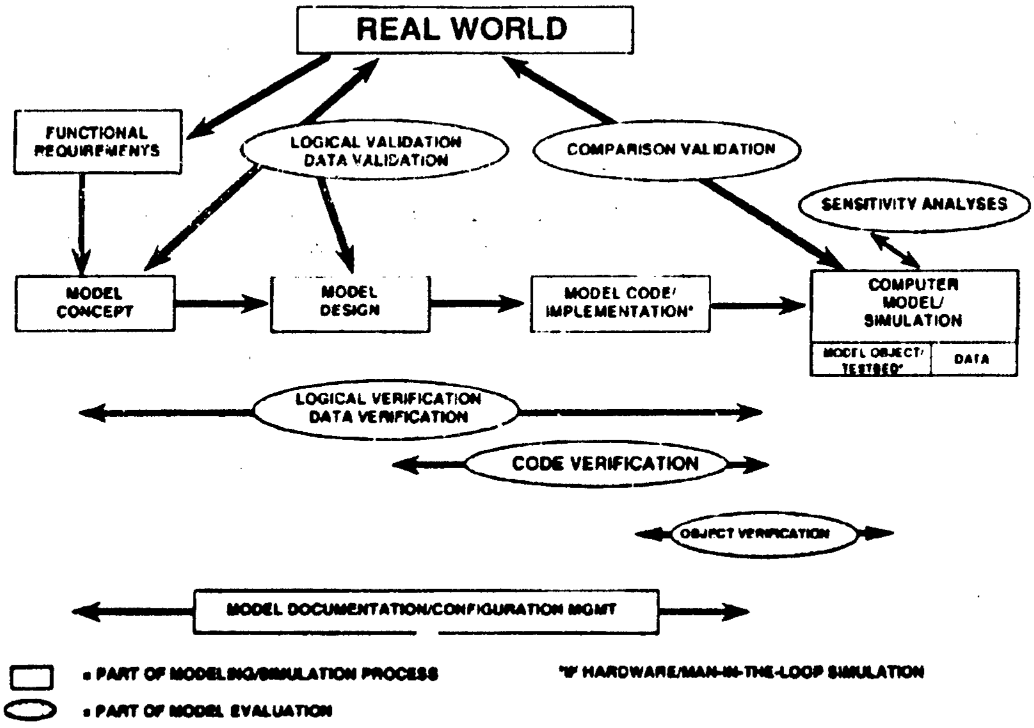

FIGURE I-4. The Relationships of V&V to the Model Form

3.0.2 Terminology

To provide a common basis for understanding, additional terminology is introduced below. Referring again to figure I-4, the development of a model or simulation involves preparing functional requirements, developing a model concept, and then proceeding through preliminary design, detailed design, and (possibly) pseudo-code to the objective computer model.

- Functional requirements. This is a statement of user requirements for what is to be represented in the model. It is an extract of the relevant portions of the real world. Ideally, it is documented in a written report.

- Model concept. This is a statement of the content and internal relationships of what is to be represented in the model. It represents the developer's concept, includes logic and algorithms, explicitly recognizes assumptions and limitations, and is documented in a written report.

- Model design. This is a highly detailed description of the model, including descriptions of algorithms, logic and data flow, input and output data, and assumptions and limitations. It may be preceded by the intermediate step of a preliminary design drawn from the model concept. The model design is documented in a written report.

- Model code (or model implementation). This is the compilable computer code. That code may be preceded by the intermediate step of pseudo-code, a computer-readable but not yet compilable implementation of the model design. The term "model implementation" applies to the case of a hardware or human-in-the-loop simulation.

- Computer model or simulation. This is the compiled and executable version of the code, including the specific computer hardware upon which that code is implemented. This is the final model to which VV&A apply. Note that a distinction is made between the code and its implementation on specific hardware because that hardware (and associated representations of arithmetic) can affect results.

3.0.3 Sample Application

To illustrate verification methods and associated techniques and tools, a sample application will be referenced repeatedly in this chapter. This involves V&V performed by a contractor on the preprocessor for a major Department of Defense (DoD) force level model. That model accepts approximately 25 million bytes of input data for a particular scenario. The preprocessor, a fully computerized process (if not a model in the purest sense), had been built and expanded over time by many programmers and had not been subject to configuration control procedures. The resulting preprocessor was not reliable, efficient, nor well documented; and was no longer a timely method of preparing input data for the force level model. The objectives of the V&V effort were, then, to apply V&V techniques to the preprocessor itself, to develop automated verification procedures for input data (for the preprocessor and hence for the force level model), and thereby to reduce the time to prepare input data for new scenarios.

The contracted V&V effort resulted in the decision by the Government to completely re-code the preprocessor in order to remove inefficiencies and add automated data verification features.

3.1 LOGICAL VERIFICATION

3.1.1 Method Description

Logical Verification is the process of ensuring, at each stage of development, that all assumptions and algorithms are consistent with the model concept. It should be performed by both the developer and an independent V&V agent.

3.1.2 Approach

Logical verification involves review of documents prepared during development to ensure consistency of assumptions and algorithms with what is intended per the model concept. Algorithms are examined. Variables treated explicitly are identified, as are potentially important but excluded variables. Values of constants are checked. Assumptions that must hold for the algorithms to apply are identified. Ideally, logical verification is performed in parallel with model development, thereby allowing for early identification and correction of inconsistencies. If performed after development or even after fielding, it can still be effective, although possibly more costly in time and analytical resources. The tools and techniques for logical verification are described below.

Documentation review. At whatever stage of model development logical verification is first applied, the collection and review of existing model documentation is the first step. A documentation review must be hierarchically ordered; that is, it begins with initial high-level statements of functional requirements for the model and proceeds down through model design specifications. (Code verification takes over beyond the model design documents and is addressed in paragraph 3.2.) At each level, the document under review is compared for logical consistency with its predecessors. "Logical consistency" does not mean absolute correctness. At each level, the developer will have a range of options with which to implement a required feature. None of the options may be "absolutely correct," particularly when the phenomenon or event is not fully understood. Some approaches may, however, be demonstrably incorrect from a technical perspective, such as using an algorithm that fails to implement the designer's stated intent. Most options, however, will be at least consistent with the intent (explicit or implicit) specified in predecessor documents. Note that the same considerations apply within any one document as well; here the issue is internal consistency.

Requirements accounting. This is a process that traces requirements from their earliest written form (in, for example, a functional requirements document) through all design documents to implementation in code. The intent is to ensure that all requirements have been accounted for. Each original requirement must be linked in "tree fashion" to one or more functions or features at the next development step. Conversely, each function or feature must be traceable back to a requirement. A failure in either direction is a requirements accounting discrepancy. This technique is often applied in an independent V&V process accompanying development of a new model under contract or major modification to an existing model via contract.

Design walk-through. This involves the design team discussing each aspect of the design with a group of functional experts in an interactive session. While it is a method of uncovering flaws in the design, it may also uncover flaws in the statement of requirements or in the specifications. Ideally, some form of requirements accounting will have been applied prior to the walk-through in order to provide both a structure to the walk-through and complete coverage of the issues. In any event, the design team may use any of a variety of systems engineering diagramming techniques (e.g., data flow diagrams) and other visual aids (e.g., symbolic and iconic models) in interactive briefings to explain each feature of the design (i.e., what requirement it responds to, how, and why). When other design options might be both obvious and attractive, the reasons for adopting the selected option should be stated. Often, the "why" of a choice is the most important information. First, it is the one most likely to uncover a mismatch between expectations and design. Second, it is the one most likely to uncover a mismatch between expectations and stated requirements. Consequently, to be effective, a design walk-through should be challenging but not adversarial. Both the design team and the review team must be ready to fully explain what they have documented and stated.

Flow diagrams. Flow diagramming (also called data flow diagramming, structured analysis, or occasionally process diagramming) is a technique that approaches any system or process from the perspective of the data being used or manipulated. It is particularly powerful and robust as a vehicle of communication between systems designers and functional users. In essence, flow diagramming uses a very simple set of symbols to record how data -- in whatever form -- moves through and is transformed by a process. Each level of flow diagram is itself analyzed using the same technique until either no further information is to be gained by going to a still lower level or the current level is sufficient for the purpose at hand (e.g., briefing an existing design as opposed to developing a new design). At a lower level — reflecting the planned or actual implementation of a computerized system — flow diagramming can also include flow charting, which is a technique that uses standard symbols to represent data flows, system logic, and physical data processing entities. In comparison to data flow diagramming, flow charting places far greater emphasis on the hardware and software mechanics of a system and thus is most useful in support of the model design for a system or in troubleshooting a fielded system.

Algorithm checks. This involves rigorous verification of the mathematics of an algorithm to ensure freedom from any errors in the expressions (e.g., incorrect signs, incorrect variables applied in equations, derivation errors) and to ensure that the algorithms are consistent with their stated intents. Algorithm checks are usually a part of a document review effort, but also may be performed without a more general review. In either case, they must be performed at both the design level and (if they pass that test) again at the pseudo-code level. The dual check is necessary because the mathematical expressions themselves change, and are subject to human error, when transformed from symbolic form in design documents into pseudo-code form.

Computer-Assisted Systems Engineering (CASE) tools. These are compiled application programs that can be used to analyze the source code of other programs.

These tools provide measures of program correctness and design efficiency, and can be used to assist in converting logical or conceptual process descriptions into a computer-based methodology. Included here are:

- Structured analysis tools
- System requirements analysis tools
- Flow charting tools
- Network analysis tools
- Performance models.

While the first three of these have obvious places in logical verification, the remaining two also have legitimate roles. For example, the system specifications for an interactive model, particularly a networked war game, may state the peak projected data communication loads, maximum acceptable error rates, etc. Performance models and network analysis tools can then be used to verify that the projected performance of a proposed design can meet stated requirements.

Sensitivity analysis. As a validation technique, sensitivity analysis involves executing the model with systematically varied input parameters to ensure that the model behaves as would be dictated by the real world. As a verification technique (and specifically as a logical verification technique), sensitivity analysis again involves executing the model with varied input parameters; however, here the purpose is to ensure that the model behaves as dictated by the model concept and satisfies the intent of the functional requirements. By implication, sensitivity analysis ensures that the model is properly "sensitive" to those factors that are essential to the functional requirements. Sensitivity analyses are necessary due to the complexity of many models. The problem is that it is frequently impossible to express the input-output relationships of a model in a single equation or set of simultaneous equations. Instead, most models and simulations achieve their solutions in a step-wise fashion. At each step, an intermediate outcome is determined from the set of input parameters to that step and in turn becomes an input parameter to the next step. Furthermore, an intermediate outcome may determine which step is next. Such situations give rise to complex multi-dimensional outcome distributions. Sensitivity analysis attempts to provide the reviewer with an understanding of such an outcome distribution and its relationship to a particular range of input values, without necessitating a full understanding of the internal complexities of the model.

Determining directly whether a model behaves as intended is frequently impossible, simply because there is no "intended" or expected outcome distribution to use for comparison. Instead, sensitivity analysis should be applied in a four-step process. First of all, the model is broken into components for which the outcome distributions are known, potentially breaking it down to the level of the modules corresponding to the individual algorithms. Second, each such component is analyzed to determine which input parameters should be varied, in what combinations, and over what ranges, to test the model-generated outcome distribution adequately. Third, the actual tests are performed using either direct one-on-one comparisons for deterministic components or statistical hypothesis testing for stochastic components. If the model passes those tests, the fourth step is to design and conduct a sensitivity analysis experiment that exercises the complete model over ranges of input parameter values and examines the resulting outcomes for logical consistency with their respective input values and with each other. Often, only the fourth of these steps is applied, and occasionally little planning goes into its design. Unfortunately, that fourth step by itself is an extremely weak form of verification. The basic problem is still that the form, shape, and parameters of the outcome distribution of the complete logical model remain unknown. When only the fourth step is performed, however, it becomes the verification analogue of face validation, i.e., the best that the verification agent can say is "For the specific set(s) of input parameter values that were used, nothing was seen in the output values that would discredit model results." If that statement can be combined with additional statements regarding how the experiment was designed to ensure that critical relationships were identified and adequately tested, confidence should increase but would still fall short of what could be achieved by applying all four steps.

Note also that sensitivity analysis for the purpose of logical verification should only be performed after code verification, because even a relatively minor code implementation problem or error could invalidate the findings of any logical verification test using that code.

Reverse engineering. This is a model assessment methodology that has application to both the logical verification and the logical validation methods. It is based on the fact that the capabilities, accuracy, and validity of a fully computerized model can be no better than those of the underlying logical model. Implementation considerations and techniques can at best preserve the attributes of the logical model, but may degrade them. Thus an assessment based on the logical model can provide an estimate of the high end of a model's attributes.

A model's analytical capabilities and technical validity attributes are determined by its logic and control structures and their underlying assumptions, its computational algorithms and underlying mathematical assumptions, and its data manipulation and transformation algorithms. The reverse-engineering approach breaks the logical model into its component algorithms and logic constructs and then derives those same algorithms and develops those same constructs from a zero base. As the derivations and developments proceed, each assumption necessary to each step of the process is identified and recorded.

For reverse engineering applied in logical verification, the assumptions are examined for logical consistency among themselves and with the requirements and precepts of the model concept. Inconsistencies are noted and analyzed for their implications *vis-a-vis* the intended application of the model. Ultimately, the user must be asked to decide whether those implications are severe enough to require adopting alternative assumptions and thus revising the model.

A shortened form of reverse engineering attempts to identify and analyze only a few "most critical" algorithms and logic constructs in the model. It further attempts to start at some level above the zero base. Doing so, however, presumes the presence (in the subject algorithms and constructs) of building block terms or algorithms that have received rigorous and documented zero-based assessments in the past. For the shortened form to be effective, model documentation must be comprehensive, particularly regarding algorithms and logic constructs.

3.1.3 Sample Application

Regarding V&V of the preprocessor introduced in paragraph 3.0.3, the first step was to review requirements to determine the logic needed. This review was accomplished through several methods, including: review of design documents, analysts' manuals, and programmer comments within the source code; interviews with users and subject area experts; and (in limited cases) review of DoD policies regarding doctrine, procedures, and reporting. Following a thorough and extensive requirements review, a new detailed system design plan was produced that thoroughly captured every portion of the logic, data manipulations, and algorithms to be included in the preprocessor. This document was reviewed by users, analysts, data source specialists, and subject area experts as appropriate. The design document was subsequently used as the basis for re-coding the preprocessor.

3.2 CODE VERIFICATION

3.2.1 Method Description

Code verification is a rigorous audit of all code (pseudo-code and/or compilable code) to ensure proper implementation of the model design.

3.2.2 Approach

Listed here are code verification techniques. Explanations are provided where meanings are not evident and have not been provided earlier in this chapter.

- Documentation review.
- Code walk-through.
- CASE tools.
- Automated test tools. A number of computerized tools exist for generating test cases, test data, and coverage measures for employed tests.
- Peer review.
- Sensitivity analysis.
- Requirements accounting.

3.2.3 Sample Application

For the preprocessor case study introduced above, code verification was applied to the original preprocessor and to the re-coded preprocessor. For the original preprocessor, code verification was performed to identify correctly and incorrectly implemented steps. For the re-coded preprocessor, code verification was performed to ensure that all design functions were correctly implemented. For both the original and the re-coded preprocessor, code verification involved thorough exercising of all functions of the preprocessor to test input-to-output relationships and processing.

The primary tools used were computer compiler and system environment tools. The primary technique applied was code walk-through. This involved individual team members stepping through other programmers' code to ensure "sanity," adherence to established coding conventions, appropriate commenting, proper file input and output control, and correct variable naming and usage. All computer source code and object code were subject to rigorous test and review by the contract development team, as well as by the end user organization. A test plan and associated documentation (such as program specifications and maintenance manual) were also prepared and delivered with the new computer code.

3.3 DATA VERIFICATION

3.3.1 Method Description

Data verification is the process of ensuring that source data that are to be used in the model are converted correctly to model input data and are consistent with the concept and logical design of the model.

3.3.2 Approach

Listed here are techniques for data verification.

- Documentation review.
- Checks of range and dimension of data.
- Plots of data.

3.3.3 Sample Application

Returning to the preprocessor case study, all input data had to be verified as correct from external sources. This step was accomplished as a cooperative effort of the contractor performing V&V and the DoD user organization. The latter has corporate knowledge of the sources and proper format of the raw data. All data computations stated in the design plan were also verified to ensure consistency and proper manipulation. This was especially true for data on various classes of supply that are to be distributed across theater regions and time periods in accordance with military populations.

Data formats are particularly important since the force level model itself applies strict conventions in its input read statements; data outside proper ranges can cause immediate read errors or, worse, model execution with embedded errors. To ensure proper format for preprocessor output data (force level model input data), rigid checking procedures were included in the re-coded preprocessor. Unacceptable output data (e.g., values outside specified ranges) would generate appropriate messages. The re-coded preprocessor was then subjected to vigorous sensitivity testing to guarantee that erroneous output data formats could not occur without warning, and thereby to ensure that data values outside acceptable ranges could not be entered into the force level model.

Additional features (appropriate to code verification or data verification) of the re-coded preprocessor included two reports: an audit trail report, and a report generator. The audit trail report is generated automatically by the preprocessor during each major processing step; and includes information on the number of data records processed from various input files, identification of records containing out-of-range values, and cross data file validity checks of unit hierarchy. This report provides statistics on program execution, as well as automated verification checks. The report generator allows users and analysts to query intermediate or final data bases for conditions or quantities. This permits further data verification. Unit organization structures, dependencies, and equipment holdings can also be reported for checking against the scenario description.

3.4 SPECIFIC LOGIC AND/OR ASSUMPTION COMPARISON

3.4.1 Method Description

This is a process of ensuring that the logic design, implicit assumptions, and explicit assumptions are consistent with a specific type of application. The process identifies strengths and weaknesses of the model for the specific application. The process is applied when a model is proposed for a specific type of application for which accreditation has not previously been granted. (Refer to Chapter 5 for further discussion on the topic of accreditation.)

3.4.2 Approach

The techniques for logical verification, code verification, and data verification listed in previous sections apply here as well. The most likely techniques to be used are the following:

- Documentation review.
- Algorithm checks.
- Sensitivity analysis.
- Peer review.

3.4.3 Sample Applications

One application of logic/assumption comparison is the user survey performed by the Warrior Preparation Center (WPC). After a war game exercise supported by WPC's family of models, WPC solicits feedback from participants on the utility of the exercise. Comments on the fidelity of the game and on its adequacy for desired training purposes are possible.

3.5 SUMMARY

3.5.1 Utility

Verification ensures that the implementation of the model accurately represents the developer's description and specifications. Short of exercising the model, verification provides the foundation to ensure that the model meets user needs. For this reason, verification can provide the basis for an initial accreditation. Returning one last time to the preprocessor case study, re-coding provided increased credibility of preprocessor data sources, processing, and output formats; and thereby increased the credibility of input data for the force level model itself. This, in turn, allowed for better identification of problem areas in the force level model. Thus, V&V applied to the preprocessor resulted in improved capability to perform V&V on the force level model.

3.5.2 Strengths, Limitations, and Lessons Learned

The advantage that verification methods have over most of those of validation is that they do not require data from the "real world" or from other models/simulations. Thus even when no comparative data are available, verification can increase the credibility of a model. Verification, however, can never be a substitute for validation. On the other hand, a comprehensive validation cannot substitute for verification. For instance, a model might produce predictions consistent with real world data, but logical verification might show that the model does not respond adequately to the original requirements. As the size of a model grows, line-by-line verification of code becomes less practical, and validation techniques become more essential.

The major limitation that applies to several of the logical verification techniques discussed above is that applying them may be as much art as science. This is especially true of reverse engineering, design walk-throughs, and documentation reviews. Furthermore, reverse engineering, sensitivity analysis, and algorithm checks require considerable expertise in mathematics, including areas such as design of experiments and statistical hypothesis testing.

To facilitate logical verification, design documentation should include detailed descriptions of the algorithms and flow charts showing how input variables are transformed (through intermediate variables) into output variables. In addition, the documentation should describe the model concept, the major processes in the model, and how those processes interact.

Significantly, documentation is often lacking, inadequate, or dated. Models are continually updated to correct errors, improve performance, or add new capabilities. Documentation lags. Where documentation is inadequate, reviewers may be forced to resort to code verification techniques -- a daunting task for a large model. Standards for documentation must be included under configuration management to ensure that logical verification can be performed.

Where an independent V&V agent performs verification, effective dialogue must be maintained between that agent and the model developers. Questions arising from the documentation (or its absence) can frequently be answered by the developers. The exchange of information and ideas assists the reviewers in correctly interpreting the documentation and understanding the model. Draft review documentation should be provided to the developers to permit correction of factual errors prior to final publication and presentation of findings.

For a time-stepped model, the review should examine the choice of the time step, which can be important for fidelity and run time. Choosing too large a step could render representation of some essential activities impossible. For example, for a combat model that represents air-to-air engagements, a one-minute time step may be too large, since air targets may enter and leave the launch-acceptability region within that time step. On the other hand, choosing too small a time step could elevate run time to an unacceptable level; e.g., for a training simulation, run times must generally be maintained at (or faster than) real time.

For an event-stepped model, the review should examine logical flow to ensure that event interactions are properly considered. Sometimes a particular event is initiated and scheduled for later completion regardless of other events that could intervene and affect its completion.

3.5.3 Life Cycle Management

Verification (indeed, V&V in general) should be seen as a continual process that parallels development and enhancement of a model. Generally, a model is developed, adjusted, and expanded over its life. At appropriate times, verification should be applied to ensure credibility of the model for its original or newly intended applications. Clearly, V&V must be included in overall configuration management, as discussed previously in Chapter II.

BIBLIOGRAPHY

Guidelines for Model Evaluation, PAD-79-17, U.S. General Accounting Office, January 1979.

Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis, Saul I. Gass, *Operations Research Vol. 31, No. 4, July-August 1983*, Operations Research Society of America, 1983.

Technical Validity Review of Lethal Area Ratio Battle Damage Assessment Algorithms, R. W. Eberth, Information Spectrum, Inc., April 1986.

DoD Simulations: Improved Assessment Procedures Would Increase the Credibility of Results, GAO/PEMD-88-3, U.S. General Accounting Office, December 1987.

Verification Revisited -- 1983, Clayton Thomas, Chapter 13, *Military Modeling, 2nd Edition*, Military Operations Research Society, 1989.

Policy Guidelines for the Application of Modeling and Simulation in Operational Test and Evaluation, Director, Operational Test and Evaluation, Office of the Secretary of Defense, January 24, 1989.

Memorandum, Deputy Under Secretary of the Army (Operations Research), October 30, 1989, subject: Verification and Validation (V&V) and Accreditation of Models.

Numerous model review documents prepared by the Center for Naval Analyses, the most recent of which are:

-- Formal review of the Cruise Missile Effectiveness and Survivability Simulation (CMESS), January 1991.
-- Review of the Backfire/AS-4 Targeting Model, March 1991.
-- Review of the ASW Model SIM II: Tactics and Kinematics, April 1991.
-- Evaluation of Naval Capabilities Assessment System (NCAS), May 1991.

CHAPTER IV

PART A - VALIDATING MODELS AND SIMULATIONS

by Donald Giadrosich

> VALIDATION -- The process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model.

4-A.0 INTRODUCTION

This chapter provides broad guidance for validating models and simulations, including helpful information for developing a detailed model validation plan, conducting an appropriate model validation, and communicating the results of such activities to officials responsible for model accreditation. Because model uses cover a wide range of purposes, complexities, and activities, the degree to which validation can be achieved in a practical manner for each model use will vary. Moreover, since many models are improved as they are used, validation of a model should be a continuous process, conducted throughout its life cycle.

Model validation can be distinguished from model verification in that verification compares the model against its design specifications, whereas validation compares the model against the real world. The term "real world" is used herein to characterize actual objects or situations, or our best representation of them. Model validation can be distinguished from model accreditation in that validation is a comparison process whereas accreditation is a decision to use a model based on some level of verification and validation.

The Military Operations Research Society (MORS) definition of model validation shown above incorporates several operative words which are extremely important. First, validation is a process; this implies it must be systematic, traceable, and describable, and the results repeatable. Second, it establishes the degree to which a model is an accurate representation of the real world. If the specifications to which a model has been developed accurately reflect the real world, the processes of verification and validation should essentially yield the same results. However, there may be important differences between the model and the real world -- some intentional and some not intentional. The validation process formally identifies and establishes the degree of the important differences. Finally, validation is accomplished from the perspective of the intended uses of the model.

Embedded in the validation process is the implied responsibility to identify and document both the proper use and the potential misuse of a model. Ultimately, for each model application, validation is accomplished and documented for the specific classes of objects (e.g., scenario(s), mission(s), weapon systems, etc.), specific levels of investigation (e.g., end game, platform performance, campaign, etc.), specific inputs and conditions (e.g., parameters, data bases, etc.), and the specific outputs of interest. In military modeling, the outputs of interest derived from the models are often described in terms of measures of effectiveness (MOEs) and measures of performance (MOPs).[1]

Because of the broad, all-encompassing definition of models and simulations, the specifics of each model validation effort must be tailored to the given problem the model is being used to solve, technical situation, or operational application. For example, if a model is being used to address the effects of flare intensities and flare drop patterns on the tracking and guidance of a given missile, it would likely be required by the accreditor that a high level of engineering validation of the flares and the missile tracking capabilities be accomplished for the model. On the other hand, a model might be accredited to investigate the probable damage that could be inflicted against a military air base by multiple attacking aircraft even though it has minimal detailed engineering fidelity regarding the specific effects of flares. This could occur if the known effects of the flares as estimated by physical testing (or a more detailed engineering model) were available and could be properly input to the air base attack model. Consequently, model validation can be limited in scope. Although tailoring for the specifics of the problem is required, the basic comparative framework put forth in this chapter is generally applicable for all types of models.

Model inputs, outputs, and internal functions all vary in the degree of accuracy with which they represent or describe the known or agreed upon state of nature. The model validation process systematically identifies and documents this degree of accuracy. Since validation is a comparison, a model can be considered sufficiently valid, i.e., "good enough," if the results of the comparison (including strengths, limitations, and assumptions) are acceptable in the eyes of the accreditor. (For further discussion of Accreditation, see Chapter 5.)

4-A.1 THE MODEL CONCEPT

The primary purpose of a model or simulation is to provide a representation of a system or relationships which can then be used to investigate the basic functions of that system or relationships. A model, in a broad sense, allows the abstraction of a study of a problem in such a way that (1) the fundamental processes of the problem and their influences and relationships can be better understood, (2) predictions or extrapolations from the outcomes of current problem conditions to potential outcomes of future problem conditions can be made, and (3) relative comparisons of alternative systems or solutions in meeting stated goals and objectives can be made.

Modern modeling has been extended to the examination of extremely complicated problems ranging from military war games to vastly complex systems like those proposed for ballistic missile defense. Modeling has been defined as encompassing "...the development of axiomatic systems, the formulation of social theories, the derivation of physical first principles, and the drafting of laws. It is thus an art natural to mankind, and focusing this art on the domain of military science conceptually encompasses the principles of war, strategy, tactics, the laws of warfare, and the structure of military forces."[2]

Scientists and engineers employ models as a means of mathematically or logically expressing the relationships between variables. **Figure IV-A-1** is a simplistic representation of this process.

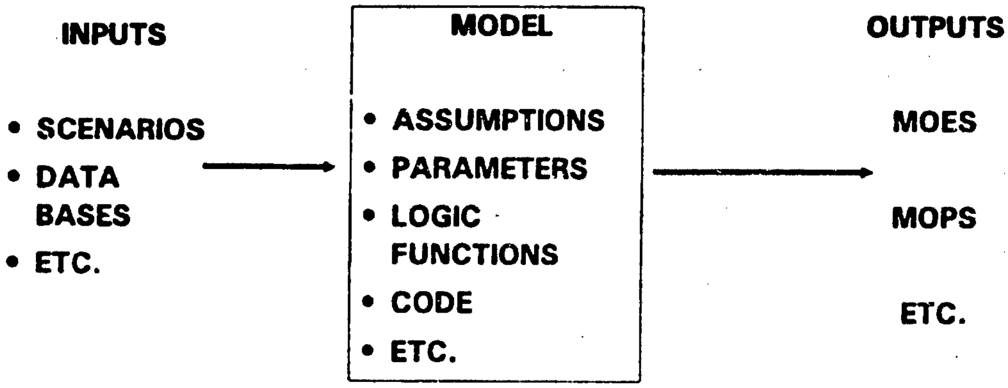

FIGURE IV-A-1. Simplistic Representation of the Modeling Process

The simple model depicted in **Figure IV-A-1** can be thought of as somewhat analogous to a scientific hypothesis based on *a priori* knowledge which is accepted as correct and from which inferences can be drawn. The model may contain axiomatic systems, social theories, physical first principles and laws, and always requires certain assumptions. When certain input data and conditions (e.g., scenarios, data bases, etc.) are provided to the model, it operates on the inputs to produce certain outputs that can be described in terms of desired MOEs, MOPs, etc. The form of the model may be analog, digital, hybrid, man-in-the-loop, hardware-in-the-loop, or some other variant; and it may be deterministic, probabilistic, or a combination of both.

A model or simulation often takes on a hierarchical structure for application to very complex problems. This structure may take the form of very detailed model which addresses each of the fundamental processes of a problem. The outputs from these models are used to provide input to the next level of the hierarchy which may treat several of these fundamental processes and their influences and relationships to each other. The outputs from this level of the hierarchy feeds the next level, etc. Each level of the hierarchy addresses a larger problem but usually in more general and less specific terms. This facilitates treating extremely complex problems in a more structured, rigorous and transparent manner than would otherwise be possible.

Modelers and simulators sometimes describe their applications of this hierarchical approach to modeling in terms of levels of analysis. Over the past decade, some analytical communities have adopted four levels of model analysis, which are defined in "A Methodology for the Test and Evaluation of Command, Control, Communications, and Intelligence (C3I) Systems" (draft), published by the Deputy Director, Defense Research and Engineering (Test and Evaluation), as part of an Implementation Program Plan (IPP) on "The Use of Modeling and Simulation to Support the Test and Evaluation (T&E) of Command, Control, Communications, and Intelligence (C3I)

Systems," dated 11 April 1990. **Table IV-A-1** illustrates these levels for air combat.

A level I model, for example, by definition "examines the performance of an individual engineering subsystem or technique in the presence of a single threat." The level I model requires as input the engineering parameters and characteristics of the subsystem in sufficient detail to ascertain the actual end game effects of the electronic countermeasures (ECM), chaff, flares, and/or maneuvers employed by a single aircraft on the performance of the system or missile. These effects can be determined internal to the model and are not required inputs. By definition, "level I outputs can be combined and fed into level II analyses to evaluate the installed, aggregate performance of a number of specific engineering subsystems or techniques against a specific threat." This definition implies that the level II model has the appropriate logic to treat the effects of ECM, chaff, flares, and/or maneuvers of the individual systems and how they combine and behave within a specific threat or threats. Levels III and IV are defined to aggregate at higher orders of complexity and conflict.

The relationship of a model or simulation to a hierarchy of models and simulations is an important consideration in validation. A model could very well be validated for stand-alone use but not be validated for use in a particular M/S hierarchy. Conversely, it may be validated for use in a particular hierarchy and not be validated for certain stand-alone usage.

TABLE IV-A-1. Example Levels of Model Analysis

LEVEL	SHORT NAME	KEY PROCESS	MODELING OUTPUT
I	Engineering	Electromagnetic Signal Flow	Electronic Combat (EC) Performance
II	Platform	Weapon System Engagement Performance	Weapon System Performance
III	Mission	Multi-Weapon System Operations	Weapon System Effectiveness
IV	Theater/Campaign	Force-On-Force Targeting	Force Effectiveness

4-A.2 DECOMPOSITION OF A MODEL

Models and simula'ꜜns, regardless of the level of complexity, can be 'hought of in terms of a number of interrelated component parts which function together to take the input information and operate on it according to specific model functions. Model functions encompass all things internal to the model (e.g., assumptions, logic, algorithms, parameters, coding, etc.). During design, an attempt is made to structure the model functions in an optimum manner to produce the desired outputs. Decomposition of a model can be thought of as a reversal of the initial design process.

Model validation is greatly facilitated by decomposing models into their component parts and subparts. Decomposition of a model is analogous to the application of engineering systems theory, whereby a total system is broken down into a number of subsystems described by transfer functions. In fact, engineering systems theory can often be employed in this process. Each subsystem is then characterized by its own model functions and associated inputs and resulting outputs. Full system examination is facilitated because a large intractable problem is broken up into a number of smaller manageable problems.

The decomposed model can be examined in terms of model functions as depicted in **Figure IV-A-2.**

Decomposition also allows the modeler to view and better understand the operation of the internal workings of the of the system. For example, our model might be the fly-out model of a specific missile. Our first level of decomposition could be examining the model function for the pitch steering channel of the missile. Decomposition

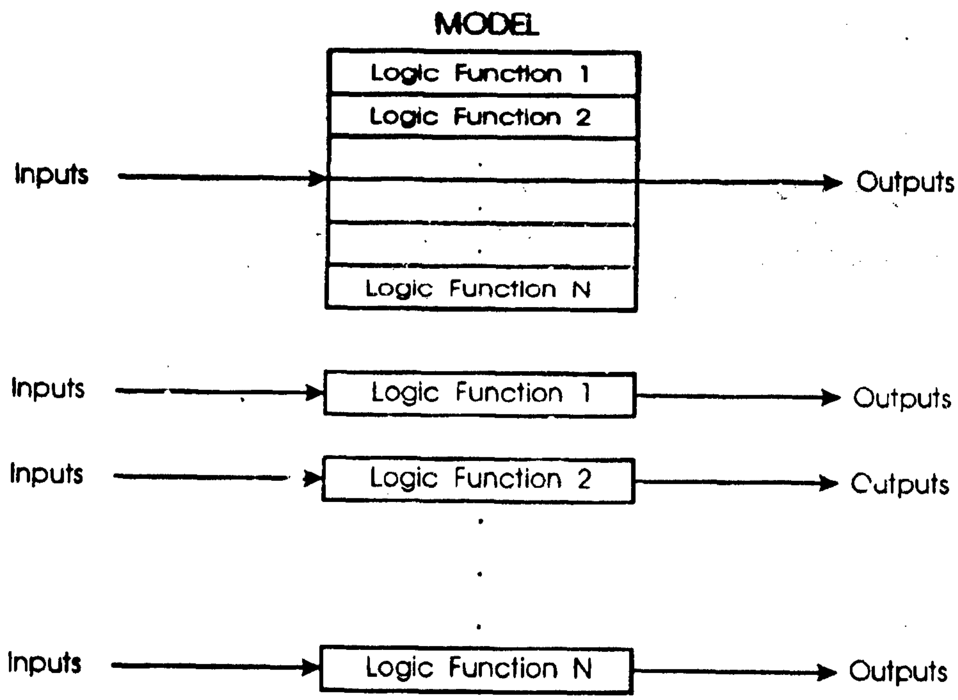

FIGURE IV-A.2. Decomposition of the Model

might be achieved to lower levels as depicted in **Figure IV-A-3**. Here, we take the specific model functions and break them down into components. The model function for missile pitch steering might be further decomposed into seeker unit, processor, torque converters, and so forth. Neither of these two graphic depictions is intended to imply that certain parallel and/or serial relationships within the model not be accounted for (e.g., the adequacy and/or completeness of defining subsystem interdependency). In fact, the components) and their interactions within the model must account for subsystem interdependency to achieve the appropriate output.

4-A.3 THE VALIDATION PROBLEM

Validation has often been used in the broadest sense as a measure of how much credence one should place in the output of a given model. We are using the term more narrowly to refer to the process of checking out the model against real world information. From this perspective, validation is part of determining the credibility of the model for a particular set of uses, not the totality of all possible uses. The validation process is designed to increase knowledge about how well the model represents reality and to aid users and decision makers in determining whether the results obtained from the model sufficiently represent what

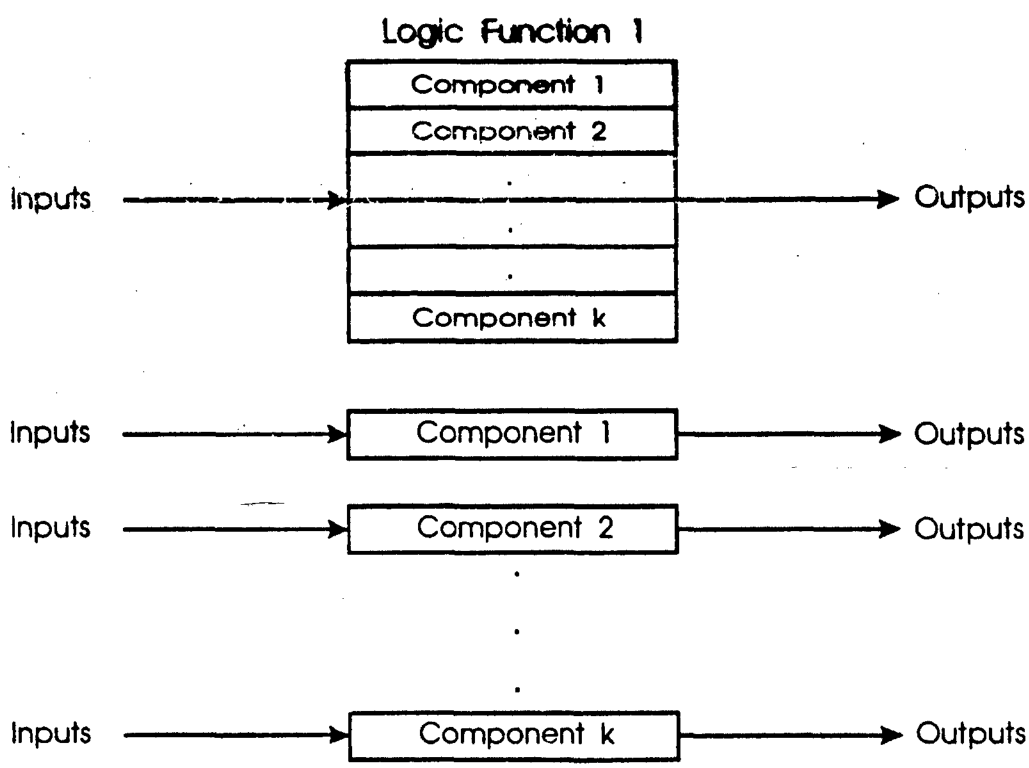

FIGURE IV-A-3. Decomposition of Model Functions

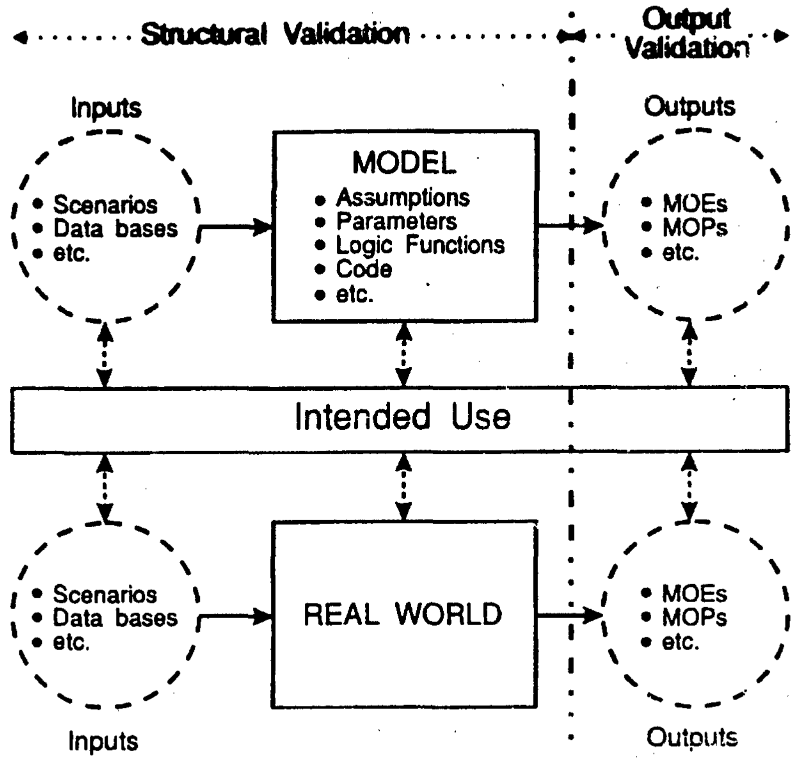

FIGURE IV-A-4. Potential Comparisons for Validation

they would observe if the situation or entity were actually played out in the real world.

There are numerous dimensions by which one can partition the comparisons that can be made in validating a model. The spectrum of potential comparisons includes elements of the inputs, the outputs, and the model itself. As illustrated in **Figure IV-A-4**, we have partitioned the outputs of the model into a domain called output validation and the model inputs along with the model internal functions into a domain called structural validation. Validation of complex models should include some combination of both structural and output validation.

Output validation is the most credible form of validation and should be conducted at the full model level to the extent possible. When it is not possible to conduct output validation for the full model, the model can be decomposed and output validation accomplished for parts of the model to the extent practical.

Structural validation should also be accomplished for those aspects of the model critical to the model's use. The planned application of the model should always be a key driver in establishing the details of its specific validation.

Output Validation

In general, output validation usually contributes the most convincing evidence for establishing the credibility of a model. It is the process of determining the extent to which the output (outcomes or outcome distributions for the model and/or sub-models) represent the significant and salient features of the real world systems, events, and scenarios it is supposed to represent. Output validation involves collecting real world data and comparing them with the output of the model (i.e., MOEs and MOPs of interest) to assess how well the model results reflect those of the "real world," i.e., the actual system or process being modeled.

In extremely complex and difficult modeling situations, the requirements for comparing real world results and model results may be difficult, if not impossible, to meet. This difficulty usually arises from the inability to actually conduct a realistic exercise of the system being modeled because of certain constraints (e.g., insufficient environmental conditions, resource constraints, safety considerations, insufficient threat representation, inability to replicate the real world conditions, etc.). Even though comparison at the output level of the full model may not be possible, it is often possible to make comparisons at lesser levels or with a scaled-down version of the system. Ironically enough, it is this inability to replicate (or even to understand) the real world that usually drives one to the use of a model in the first place. Quantitative approaches to the comparisons usually provide the most convincing evidence about model validation. Quantitative output validation, however, requires that the outputs of the model be observable and measurable in the real world.

Qualitative assessments made by operational and technical subject matter experts are important inputs to the validation process. As discussed earlier, under certain conditions, quantitative comparisons may be prohibitive or limited. Face validation by experts and other qualitative methods (e.g., the use of focused-group interviews or a modified Delphi technique) for obtaining expert opinions on critical model issues should be applied. Findings from the qualitative methods should be used to supplement and reinforce the available quantitative comparisons.

Model outputs are generally selected based on how well they represent the military performance and utility of a system (i.e., MOEs and MOPs as discussed earlier). MOEs and MOPs represent different sets of system measures of interest from the perspective of operators and developers, respectively. As depicted in **Figure IV-A.5**, these two sets are not necessarily mutually exclusive. For example, it is highly probable and desirable that both operators and developers have a keen interest in some of the same measures for certain systems and situations. Also, some form of functional relationship normally exists between the MOEs and MOPs of interest, even though they may not be well-defined or explicit.

Furthermore, it is extremely important that observations and measurements made in the real world be executed in such a way that they accurately represent the outputs of interest. This implies the application of the scientific approach to testing and experimenting and the inclusion of quantitative and qualitative statistical comparisons where appropriate. Such comparisons may be made based on data points, intervals, and distributions and may involve both absolute and relative values. Data from testing should be model-compatible to the extent possible so that the model-test-model approach to development can be invoked.

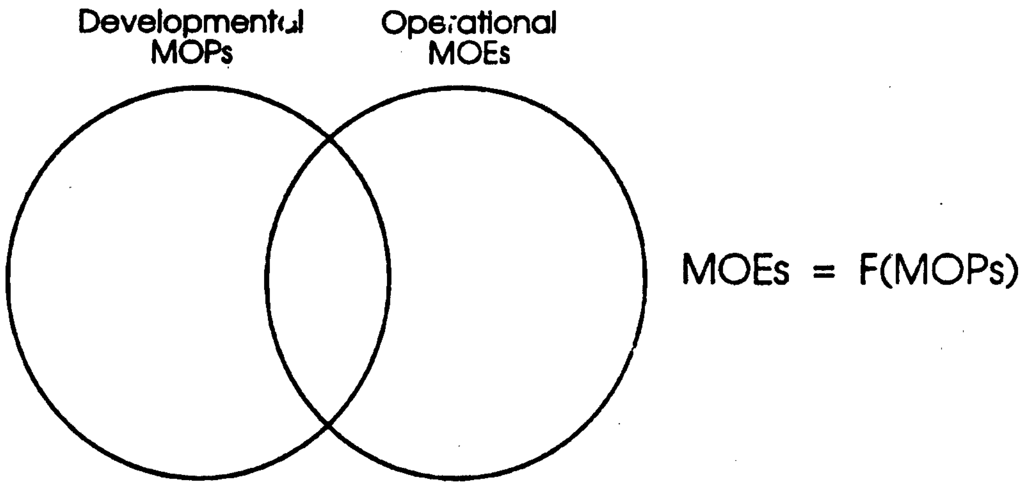

FIGURE IV-A-5. Output MOEs and MOPs

4-A.3.1. Structural Validation

Structural validation deals with an *a priori* examination of the model input data, the basic principles of the model, and its assumptions to determine the degree to which they are complete, logically consistent, and reasonable for the types of uses envisioned. To a large extent, structural validation is designed to increase the knowledge and confidence of developers, users, customers, decision makers, and independent reviewers in the model results by demonstrating that the model has internal integrity. Structural validation is the process of determining the extent to which the input data (e.g., scenario(s), mission(s), data bases, etc.) and the model (e.g., assumptions, logic, algorithms, parameters, coding, etc.) represent the significant and salient features of real world systems, events, and scenarios. Model decomposition greatly aids in this process.

Structural validation may be the primary form of validation that can be accomplished with extremely complex models, especially when observations and measurements of the outputs of interest are not possible or practical in the real world. Structural validation can occur through empirical measurement and comparison, as well as theoretical examination based on physical first principles, logic and axiomatic systems, and other scientific laws. Such comparisons can include qualitative approaches, such as expert opinions, or statistical tests based on both quantitative and qualitative data. As one proceeds to greater refinement and depth in structural validation, several dimensions on which to base these comparisons may be examined. For example, one might pursue validating critical parts of the model structure in terms of theory, performance against other accredited models, scaled down or limited component testing, and any available combat history relating to the specific portion of the model being addressed.

Table 4-A-2. Examples of Information Sources

•	Functional Experts
•	Operational Experts
•	Scientific Theory (physics, engineering, behavioral, etc.)
•	System Design Information
•	Laboratory Measurements (components, response functions, etc.)
•	Special Test & Training Facility Measurements (hardware-in-the-loop, antenna patterns, radar cross section, infrared, millimeter wave, human factors, decision making, etc.)
•	System Level Measurements •• Developmental Test and Evaluation •• Operational Test and Evaluation
•	Live Fire Test Measurements
•	Combat History and Measurements
•	Other Accredited Models

4-A.3.2. Sources of Information

Model validation requires the use of a broad range of sources of information (see Table IV-A-2). This information is both qualitative and quantitative and can vary from the opinion of functional experts, operational experts, and scientists to comparisons based on precise deterministic and probabilistic measurements. Scientific theory, system design information, and laboratory measurements can provide essential structural information. Special test and training facility measurements, as well as system level measurements during developmental and operational test and evaluation, are excellent sources. Finally, live fire test measurements, historical information and combat measurements, and data (which include both pertinent information from VV&A and relevant output data) from other accredited models used for similar applications are potential sources of validation information.

4-A.3.3. Sensitivity Analysis

The sensitivity analysis is a critical part of both structural and output validation. Sensitivity analysis is a formal examination of how output variables of the model respond to changes in inputs, assumptions, parameters, and critical logic functions. Sensitivity analysis should be conducted on the total model and on all its decomposed parts. Sensitivity analysis can be used to check for proper responses to input variables and to identify marginal break points and special limiting values. It can be used to understand better how the model works and to help identify errors in the model structure and/or code. However, sensitivity analysis is limited to telling you only what the model is sensitive to; one has to go further with the comparative validation process to ensure that the model sensitivities are indeed representative of the real world.

A vital part of sensitivity analysis is to help one understand where the model results are extremely sensitive to changes in model algorithms, input data, parameters, and/or assumptions. These sensitivities should be of paramount interest to those who have to validate and accredit a model as well as those who must rely heavily on the model output for important decisions. This discussion addresses the situation where a change in a given part of the model is found to not have a significant or pronounced change on the major output of interest. For example, it might be that doubling the reliability of a given subsystem will only slightly change the overall total mission reliability because that particular subsystem does not really make a difference in being able to successfully complete the mission.

4-A.3.4. Tasks for Model Validation

Example tasks that should be accomplished and the results documented during model validation can be discussed in terms of preparing to conduct the validation, conducting structural validation, and conducting output validation. These specific tasks are listed in **Table IV-A-3**.

During preparation for model validation, it is critical to specifically define the problem to be modeled and addressed (that is, the problem or class of problems for which the model is being validated). Definition of the problem will, to a large extent, establish the intended application of the model. It is important to remember that model validation is accomplished for a particular problem (application) or class of problems. One must develop and/or select the appropriate scenario(s) and mission(s) to address the problem and assess their realism in terms of the real world. The MOEs and MOPs required to address the specifics of the problem will be the primary model outputs examined. It is important at this stage to take into account whether or not this model will be used in a stand-alone role or as one level in a model hierarchy. This will impact the inputs and outputs and how model realism needs to be addressed.

Once the above tasks are completed, output and structural validation (required to produce data to address the MOEs and MOPs) can be addressed. Generally, the amount of output validation that practically can be accomplished will influence the amount of structural validation necessary. Quantitative output validation normally is performed as a comparative test or experiment which provides a quantitative assessment of the agreement of the model with the real world. Sensitivities of the model output MOEs and MOPs to inputs, critical model logic, and assumptions should be identified and quantified to the extent practical. When output validation cannot be performed at the full model level, the model should be decomposed and structural validation applied. As discussed earlier, this involves a comparison of input parameters, data bases, assumptions, and model functions with the real world. Sensitivity analysis should always be a key tool during both structural and output validation.

4-A.3.5. Stakeholders In Model Validation

The model developer, user, independent reviewer, customer, and decision maker all have an interest and responsibility in the validation of a model. If a new model is being developed for a given use or class of uses, the model developer should verify and validate the model to the extent practical and document those results. (Verification, as discussed in Chapter 3, should be performed

Table 4-A.3. Example Tasks for Conducting Model Validation

CATEGORY	TASKS
Preparation	• Define specific function/system to be modeled and addressed. • Develop/select level of model required (end game, one-on-one, campaign, etc.). • Develop/select scenario(s), mission(s), etc. (address reasonableness versus real world). • Determine whether model will be used as stand-alone or as one level in a hierarchy. • Identify specific model output MOEs and MOPs required to address the problem. • Identify input from and output to other models. • Select and implement the appropriate category or categories of validation.
Output Validation	• Quantify agreement of output MOEs and MOPs of interest versus real world. • Quantify sensitivities of model output MOEs and MOPs of interest to inputs, critical model logic, assumptions, and parameters. • Conduct face validation and other appropriate forms of expert qualitative assessments. • Compare input scenarios, parameters, and data bases versus real world. • Address adequacy of inputs from other models and outputs to other models. • Address assumptions versus real world.
Structural Validation	• Address total and decomposed model functions versus real world. • Address sensitivities of model output MOEs and MOPs of interest to inputs, critical model logic, assumptions, and parameters. • Address interdependency of logic functions. • Address adequacy/completeness of model logic. • Address Adequacy of model in context of model hierarchy.

routinely as part of the programming and checkout phases of a model's development.) In the early stages, the validation effort likely will be more structural- than output-oriented.[3] When a user other than the developer selects and applies a model, the user inherits a responsibility for properly applying the model as well as conducting any additional verification and validation necessary for the problem at hand. (The user will also want to ensure that the model or simulation has been accredited for his application, as discussed in Chapter 5.)

When critical issues are to be addressed by modeling and simulation, it is beneficial to have someone other than the developer and user (i.e., an independent reviewer) conduct additional model verification and validation. This is designed to provide a separate and objective look and, optimistically, offsets any biases that the developer and user may have. The customers and decision makers also have a high stake in model verification and validation. For example, the decision to procure a major weapon system may be based largely on model results; and the validity of the decision could depend on the validity of the model. Furthermore, those responsible for accreditation of the model will rely on verification and validation reports in arriving at their decision (see Chapter 5).

As discussed in Chapter 2, it is essential to maintain configuration control of models. When there are multiple users of a model and the various users are modifying the model to accommodate their particular needs or usage requirements, then each version will require validation (and accreditation) for that particular application.

4-A.3.6. The Model Validation Plan

Validation of any given model should be a continuing process with appropriate documentation of the results at various key application points throughout the life cycle of the model. A formal plan should be developed to conduct this validation. The validation plan could be developed sequentially over the lifetime use of the model, with a basic plan covering the initial validation and supplements as needed to address each unique application and/or configuration update. The information set forth in the validation plan should be sufficient to supplement other program and decision making documentation, as well as to serve as the road map for validation of the model at a specific point in time. The validation plan and report should be the key documentation that supports the decision to accredit the use of a model. A sample format illustrating the types of information that should be included in the validation plan is provided below.

I. EXECUTIVE SUMMARY

II. BACKGROUND
- Purpose of the Validation Effort
- General Description of the Model
- Previous and Planned Usage
- Program and Decision Making Structure

III. PROBLEM
- Specific Problem(s) Being Addressed
  - •• MOEs and MOPs
  - •• Critical Evaluation Issues
- Critical Validation Issues (Related to Critical Evaluation Issues)

IV. APPROACH
- Validation Task(s) To Be Addressed
- General Approach to Validation
  - •• Scope
  - •• Limitations
- Specific Approach to Validation for Each Task

V. DESCRIPTION OF VALIDATION EXPERIMENT
- Output
- Structural

VI. ADMINISTRATION
- Validation Management and Schedule
- Tasking and Responsibilities
- Safety
- Security
- Environmental Impact

VII. REPORTING

VIII. ATTACHMENTS (As Required)

The background information in the validation plan should sufficiently describe the model, its present configuration, and previous and planned applications. It should also relate how the model fits into the overall program and decision-making structure. The specific problem that the model is to address should be clearly delineated along with the MOE(s) and MOP(s) of interest. Critical evaluation issues related to the problem, along with the specific tasks planned for the validation process, should be addressed. Critical validation issues should be identified and addressed in terms of how they relate to the critical evaluation issues that must be addressed by the model. The general and specific approaches to be used for validation tasks should be addressed. Planned validation experiments for both output and structural validation should be described. Finally such tasks as scheduling, planning for administration and management, tasking and responsibilities, and reporting on the validation efforts should be formally documented in the plan.

4-A.3.7. Documentation of the Model Validation Efforts

Formal documentation of the model validation activities and reports for each model validation effort are essential for proper life cycle management of the model. The documentation and reports should be directed at assisting the customer and the decision maker in the model accreditation process (i.e., it should provide information that helps the decision maker decide whether the model is good enough for the specific application and problem being addressed). Documentation of prior validation efforts also assists those tasked to conduct subsequent validation efforts (i.e., it should provide the basis for accumulating validation information). The documentation efforts include collection of information and data as the validation plan is executed and analysis and reporting of the model validation results.

A sample format illustrating the types of information that are of interest in the validation report is provided below. Ideally, the executive summary will concentrate on the model, critical issues regarding its application(s), and its strengths and weaknesses for addressing the specific problem(s) in terms of the real world comparisons made during validation. Sections II through V provide background on what was planned for model validation and, except as modified, could be extracted from the plan. Sections VI, VII, and VIII address both general and specific results of the validation effort along with a detailed accounting of the specific validation findings. Model trends and sensitivities, as they relate to the problem and the critical evaluation issues, should be described. Quantitative methods are highly desirable for the validation comparisons and should be documented. Qualitative methods also are useful and, because they usually include more subjective judgments, may require even more documentation.

I. EXECUTIVE SUMMARY

II. BACKGROUND

III. PROBLEM

IV. APPROACH

V. DESCRIPTION OF VALIDATION EXPERIMENT
- Output
- Structural

VI. RESULTS OF VALIDATION EFFORTS
- General
- Validation by Task
- Discussion of Critical Issues
- Discussion of Model Trends and Sensitivities

VII. SPECIFIC VALIDATION FINDINGS

VIII. ATTACHMENTS (As Required)

4-A.3.8. Special Considerations When Validating Models

Model validation will always require some level of judgment, but to the maximum extent possible, empirical comparisons should be made. As discussed earlicr, the process of validating a model will never be exhaustive. There will always be some things not addressed during validation, and others not addressed to the degree that some would like. Consequently, for model validation to be a productive process, it must concentrate on the specific uses of the model and the actual validation issues addressed regarding those specific uses. The validation comparisons should present a reasonable, systematic examination of the model, and an objective picture of its true capabilities and limitations in that application. Both the strengths and weaknesses of the model for addressing the stated problem(s) must be communicated to the accreditor by the validation process.

The validation process should be extensive and robust enough to properly consider the findings and views of neutrals, advocates, adversaries, and other interested parties. The goal should be to communicate all important findings regarding model comparisons and critical validation issues to the accreditor. When serious competing views emerge on critical validation issues, it may be necessary to pursue further validation efforts that can provide additional objective comparisons and information for consideration by the responsible accreditation authority.

There should be test data available for comparison on each critical issue to be addressed by the model. If feasible, it is desirable to collect two sets of real world data -- one for structural comparisons and another for output comparisons. The validation process should be such that when data derived from realistic field and development testing raise questions about prior assumptions and/or propositions of the model, these questions are addressed. The process of validation must shed light on what we do and do not know about the model's structural content, its internal functions and capabilities, and its output accuracy. Our analysis of conflicting or discrepant information often provides the insights necessary for improving the models and obtaining better answers to difficult questions.

Independent technical and operational experts can examine the model processes, its assumptions, inputs, and outputs to arrive at their opinions of the appropriateness and validity of the model and its associated results. In the academic world and in the field of operations research, this independent review is often performed by a separate unbiased party (i.e., a referee) who is responsible for helping maintain the objectivity of the analysis. Unfortunately, when dealing with highly complex and often classified systems and techniques, this objectivity can be somewhat limited, especially if documentation is not adequate. Therefore, it is incumbent upon all interested parties (e.g., model developers, users, decision makers, and other responsible authorities) to ensure that the validation process is objective, comprehensive, and well documented.

4-A.4 SUMMARY

The Military Operations Research Society advocates the formal determination and documentation of model credibility through a three-part investigation involving verification, validation, and accreditation. This chapter addressed the second-part, model validation, which is defined as "the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model."

A model can be portrayed in terms of its inputs, the model itself, and its outputs. Thus, when validation comparisons are made against the real world or physical theories and laws associated with the real world, they can be addressed in terms of model inputs, the model itself, and model outputs.

Model validation has been partitioned into two parts: (1) output validation and (2) structural validation. Output validation is the most credible form of validation and consists of comparing the output of the model against real world observations. Structural validation involves determining the extent to which the input data (e.g., scenario(s), mission(s), data bases, etc.) and the model (e.g., assumptions, logic, algorithms, parameters, code execution, etc.) represent the significant and salient features of the of real world systems, events, and scenarios. Decomposition of the model into fundamental model functions and components aids in the process of structural validation.

Validation of complex models requires an appropriate combination of both structural and output validation. Maintaining configuration control and essential documentation are important. A formal plan for model validation, along with adequate reporting and documentation of the results as described herein, are vital parts of the model validation process.

CHAPTER 4

PART B - THE MULTIDIMENSIONAL SPACE OF VALIDATION

by Dale Henderson

4-B.0 INTRODUCTION

Almost any discussion of the kinds of validation and of the activities involved quickly becomes confused and confusing. One reason for this is that the "space" of validation activities has many separable dimensions. If discussants do not first describe this space they run a great risk of individually focusing on a different "coordinate" in this space; they are then discussing different aspects of the problem using much the same words. This limits agreement to the superficial and generally results in confusion.

We have decomposed the space of validation or validation activities into five component dimensions in **Figure IV-B-1**. These separate dimensions describe (1) the techniques used, (2) the basis of truth used, (3) the applications intended for the model or simulation, (4) the degree of composition of the model, and (5) the depth of the validation effort itself. We also indicate a possible sixth dimension, the verification of the software against its own standard or standards. This sixth dimension would really be found to compose several dimensions, which may largely be discussed without reference to the present five. That is, it composes a separate space which is studied in earlier sections of this monograph.

A given activity in evaluating the validity of a model or simulation will generally be a point in this multidimensional space; but, not all points are occupied by validation efforts. For instance, the activity called Face Validation is easily identified. Its depth is usually shallow or at the surface. The Delphi technique is most common. And the model applications assumed are most often analysis. The activities of Face Validation are, however, less localized along the other two axes: the degree of decomposition of the model is generally irrelevant and the basis of truth is commonly history, trial data, or theory.

4-B.1 APPLICATION

The application coordinate is meant to recognize that models are employed in qualitatively different fashion and that what is even meant by validation depends on the sort of application. We indicate three kinds of applications which form a rough progression: synthesis, analysis, and prediction. A synthetic application is the use of a model to provide meaning or consistency, to further human understanding, but without adding any new knowledge; synthesizing understood pieces into a larger whole. Training models and simulations fall into this category. An analytic application is one to further understanding by providing a structure for further abstraction; human understanding is extended through their use. Such applications could be tests of historical records of battle against purported explanations or correlations. A predictive application is just that: a prediction of something new which can then be observed for its degree of compliance. (These distinctions are well developed in the RAND report "Is It You or Your Model Talking? A Framework for Model Validation," by James S.

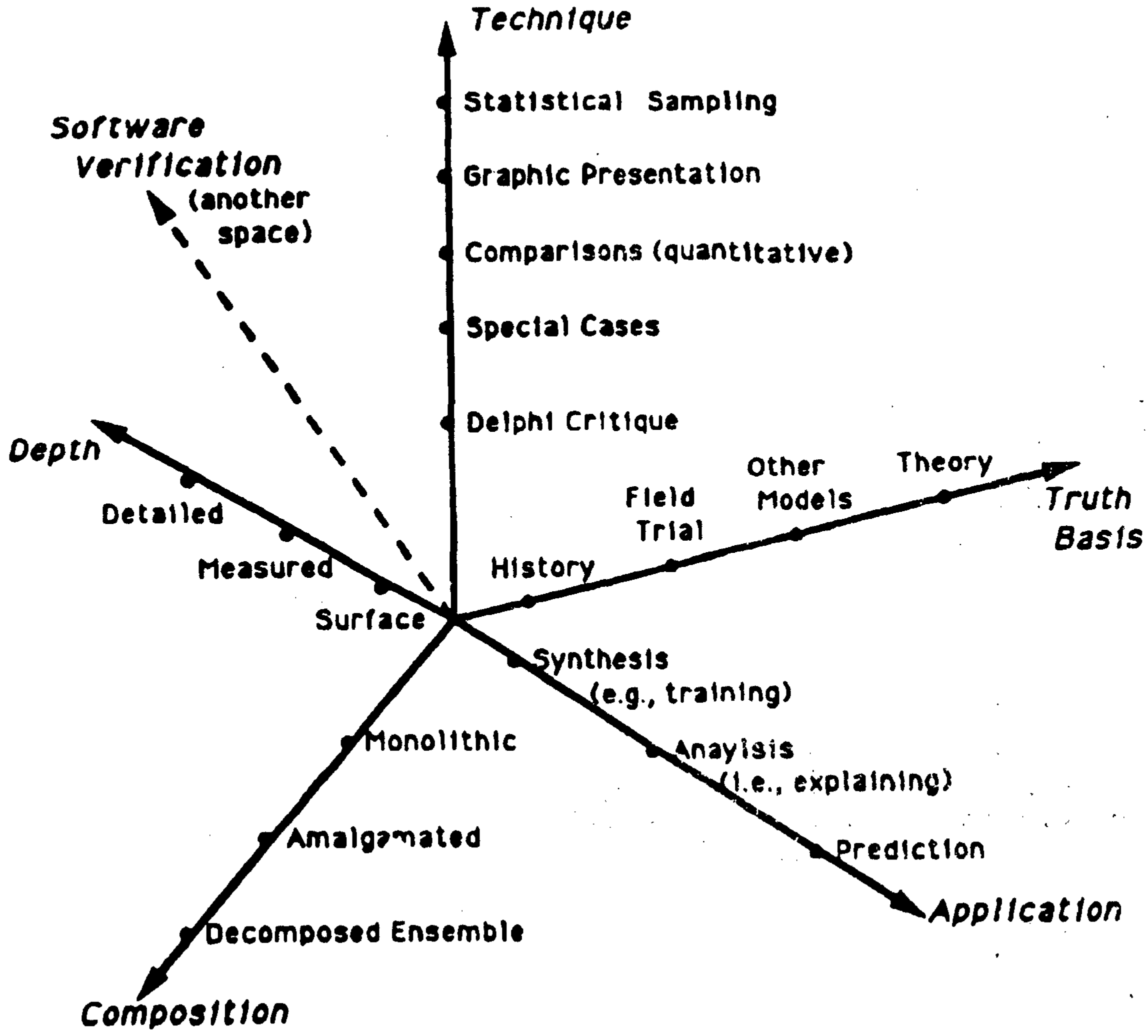

Figure IV-B-1. The Multidimensional Space of Validation

Hodges and James A. Dewar, R-4114-AF/A/OSD, 1992.)

4-B.2 TRUTH BASIS

The basis of truth coordinate represents the differing norms against which a model under consideration may be judged: "What is correct?" Obvious data include historical records (generally of combat), data from field trials, the output from other models, and *a priori* theory. No matter which bases are used, the correctness and applicability of the data should be explicitly examined. Unquestioned assumptions of applicability are especially dangerous.

4-B.3 TECHNIQUE

The several techniques are arranged in

an order indicating the amount of data considered in the process. The Delphi process consists of a group of purported topical experts functioning in some structured manner. Their analyses may skip between surface values, trends, consideration of the implications of the composition of the model, or implications of the algorithms employed. While the volume of data considered will be small, the genius of the process is that it will be the most appropriate, being somehow selected by the panel of experts.

Special cases are specified scenarios or instances for which we have some reason to believe that the correct results are known. Often these are limiting or extreme cases under which many factors become unimportant or under which algorithms mathematically simplify to analytically known results. Other special cases could include field trial results for which measurements exist. Comparisons are similar, but involve more data or trends of data.

Appeals to a truth basis from an accepted model are usually through a fairly exhaustive set of input and output comparisons. It is also common to employ graphical presentations because people can often extract (qualitatively unmeasured) trends from appropriate graphical presentations. The most efficient comparisons, to any set of truth data, employ statistical sampling, well-designed numerical experiments to explore as many sensitivities as possible with each case considered.

4-B.4 COMPOSITION

If a model represents one thing, by itself, with all external interfaces through parametric and known data, then a monolithic model is appropriate. Even without meeting these restrictions, they are commonplace, (albeit often only because of poor software design). At the other extreme are models and simulations composed of objects which are in turn composed of parts or elements. This composition is in accord with modern software practice, but raises several issues with respect to validation. First, the separate and separable *objects* can be considered separately; different techniques and truth bases being applied to different parts. But the interactions between objects must be confirmed as an additional consideration. The middle case, labeled amalgamated, is just meant to indicate the intermediate case involving a few objects, these perhaps having been abstracted from a . . her ensemble for some purpose.

4-B.5 DEPTH

The depth dimension is a measure of the degree of quantitative detail. Surface measures are just that, often applied in Face Validation by people. Formal measures are quadratures or other abstractions, perhaps correlations among the data themselves, which are of primary interest to decision makers. The detailed variables involved in a calculation (all of them) are ultimately available for comparison with (say) field trial data or with the results from other, accepted, models.

See *Appendix A* for Selected Bibliography

CHAPTER 4

PART C - FACE VALIDATION AND FACE VALIDITY

by D. P. Gaver

> FACE VALIDATION is the process of using informed experts to examine the conceptual background, execution, and output of a simulation model.

The examination begins by a review of the operational questions that the model is designed to address; thus the relevance of the model is first examined. These questions will often be quantitatively expressed ("How much of Items x, y, z should I stock at locations u, v, w for quarter 3 of 1993?"), and degree of success in answering them correctly may be expressed quantitatively, i.e., in terms of Measures of Performance (MOPs) and Measures of Effectiveness (MOEs). The experts performing face validation will first ask if answers to the questions addressed by the model will assist the client decision maker. Then the expert will comment on the way those quantitative questions are answered: are the answers comprehensible to the client, are the restrictions implicit in the modeling approach made clear so that the model will not be misused, are uncertainties in the model conclusions adequately portrayed as these depend upon model inputs (data on parameter values, etc.) and organizational structure and behavior? The experts performing this first stage of face validation are conducting an overview for model *relevance*, and *usability*.

The second state of face validation is an independent overall assessment of the credibility of detailed model output. This can be approached initially by checking whether apparently correct numerical values result in known cases, i.e., when certain parameter values are specified, e.g., set equal to zero. Obedience to physical laws can be checked, if relevant. Correspondence to other models' outputs can be checked; such other models can be simple "back of the envelope" creations of the experts themselves. The face validation experts might also ask, parenthetically, what the present model offers that an existing model does not. In the process the model's output options can be critiqued: are there informative graphics? Are tables of numbers arranged so that their implications are clear? Are numerical results expressed to credible accuracy, not to absurdly many significant digits? Are error assessments of results given in a believable and comprehensible way (documentation should cover this)? All of these steps address the overall question of how well the model does what it advertises to do.

The flavor of face validation is that the above steps are carried out relatively quickly by one or more experts in the subject-matter area covered by the model (e.g., theater level modeling, air defense, logistics, antisubmarine warfare, intelligence). The result of the face validation process can be either an endorsement of the model as is, suggestions for model revision, or outright model rejection. It is desirable that a model under development be subjected to a face validation process during the development process. It seems especially expedient that exposure to face validation processes by the ultimate client-user and his resident experts be conducted at intervals during the development process.

CHAPTER 4

PART D - SENSITIVITY STUDY OF A SIMULATION MODEL

by D. P. Gaver

The **sensitivity study** of a model has several aspects. In general such a study is made in order to check for plausibly proper model response to different levels of input variables over their jointly appropriate ranges. It is often easiest to check responses to special limiting values, at which the proper response is known from very simple, common-sense considerations or quite basic physical laws. By *response* is meant the quantifiable or classifiable outcome of an operation or experiment that the model is supposed to predict or represent. In the context of military applications a response could be weapon delivery accuracy as a function of delivering platform's speed relative to the target, maneuvering actions, and protective jamming or use of diversionary decoys by the target.

A secondary, but important, feature of sensitivity testing is to assess the degree of model response sensitivity to convenient modeling assumptions that cannot necessarily be checked for validity. For instance, weapon dispersions from aimpoints might be typically taken to be independently normal- or Gaussian-distributed; times to equipment failure might be represented as exponentially distributed random variables; and arrivals of messages at a communication center as a Poisson process. It is useful to examine responses when such assumptions are relaxed in plausible ways. *Insensitivity* of response when the strict assumptions are relaxed, given the values of, say, the mean of the corresponding distributions, is a reassuring virtue, for no simple model of a system components can be guaranteed to hold precisely

Sensitivity studies are typically conducted by comparing model results, i.e., the pattern of responses when input variable values are changed, to comparable results from other validated models, to relevant experimental or field data. The input of expert judgment can also provide valuable guidance.

A systematic and well-conducted sensitivity study will help isolate errors or omissions in the model's structural formulation as well as in the enabling computer code. It will usefully employ expertise. It will identify information and data needs and criticality.

Since many important models must represent one or more meaningful responses in terms of many input variables it can be anticipated that the use of systematic experimental design tools, such as fractional factorial and response surfaces, should and do prove useful for better understanding a model's behavior.

CHAPTER 5 - ACCREDITATION

by Ernest Seglie and Patricia Sanders

> The official determination that a model is accepted for a specific purpose is called ACCREDITATION.

5.0 INTRODUCTION

Operations Research applies the scientific method to the analysis of military operations and the utilization of military assets. Operations researchers use the powerful technique of computer modeling to form their conclusions. The British, who invented it, define OR as:

> the attack of modern science on complex problems arising in the direction and management of large systems of men, machines, materials and money in industry, business, government and defense. Its distinctive approach is to develop a scientific model of the system, incorporating measurements of factors such as chance and risk, with which to predict and compare the outcomes of alternative decisions, strategies or controls. The purpose is to help management determine its policy and actions scientifically.

Such policy and actions are usually implemented by a decision maker, who may start by knowing little about the techniques and methods of operations research and nothing about the models used during the research.

The researcher must communicate to the decision maker the results of the research and something else as well: the appropriate measure of confidence in the results. For the decision maker, to act on the results of the model research is to give credence (in some degree) to the model. The decision maker must have some reason for believing that the model is acceptable for the purpose to which it was put.

This chapter is written from the point of view of the operations researcher who must get a model accredited. For the operations researcher, the accreditation process is the way to provide what the decision maker needs in order to give the appropriate credence to the model and its use. The formal process of accreditation should mirror the processes that accompany all research: the process of convincing oneself that the methods and results are reasonable, appropriate, and worth believing to some extent, and of assessing the confidence one has in the results.

In simpler times accreditation could be done in *ad hoc* fashion. The decision maker knew the researcher and the quality of the researcher's work built on possibly years of interaction. The researcher knew the decision maker personally and could ask what was important and what was the real question. Such a close partnership characterizes some, but not most, efforts today. Often there are several layers—or filters—of management between the researcher and the decision maker. As a result, more care must be taken in the communication process.

Accreditation is, in part, a response to the gap that has developed between the operations researcher and the decision maker. If used properly, the accreditation process can reestablish the close partnership that is necessary to do relevant, and credible work. The aim of accreditation is the mutual agreement by the researcher and decision maker on the extent to which the model can be the basis of decision.

In sum, Quade and Carter note in a discussion of the "The Modeler's Versus the Decision Maker's View of Quality," that ". . .operations research has lost its most important roles because it has devolved from a market orientation based on the client's needs to a professional orientation based on tool development."[1]

5.1 DEFINITION OF ACCREDITATION

Accreditation is the official determination that a model is acceptable for a specific purpose. But what makes it different from Validation? Validation is the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model. Every model is a simplification of, and a distortion of, the real world. A determination of complete validity is therefore impossible. No validation is expected without many caveats. The best caveats (1) specify the range of values of variables over which the model has been checked (its field of validity) and then (2) specify the error that the model generates within the specified input field (degree of validity).

The degree of validity must reflect the error in the output of the model. In order to determine the error, data must be available or a test must be run against which to compare the model predictions. The result of a validation comparison might be, for example, that the model is only good to an order of magnitude, or that the model is able to calculate the range of detection in free space to within x, if the Radar Cross Section (RCS) of the target is greater than z and known to y. The validation should also address the nature of the error—whether it is systematic or random, and what aspect of the model is causing the error. Validation should discuss the data base from which the model was derived and the data base to which the model outputs were compared to determine the error.

The output from the validation may say in summary that, "When employed to evaluate force-on-force engagements of battalion size, the loss exchange ratio of a single run may differ from other single runs by a factor of three. They differ from typical training exercises by factors of from two to ten and differ from the experience of actual combat (not used in the development of the model) by factors of two to a thousand." Validation says a lot about the model from the perspective of the intended uses of the model. But validity is separate from the specific use of the model in a specific decision process.

Preparing an accreditation begins by understanding how the model outputs are to be used in the decision process. This understanding is also the starting point of the operations research itself. What is the question? What are the information needs to answer the question? Why do you need the model? What will the model produce that is important for the answer to the question? Accreditation will be the determination that the model outputs are important to the decision, and the determination that the degree to which the model represents the "real world" is sufficient for the purpose to which the model will be put in the decision process; that is, accreditation must account for the specific use.

For example, in the case of exchange ratios discussed above, if the approximate difference in cost of two alternatives is expected to be 30%, this may imply that performance differences of that order are important to know. Note that this is a question not directly addressed in any of the validity caveats mentioned above.

If the detection model above is used to explore whether holes in sensor coverage appear, such a determination will depend on the density of sensors as well as the model. That the RCS will be known only to δ may mean that the model cannot be used to explore whether or not holes open up in the sensor coverage.

Thus the accreditation process must take the demonstrated degree of agreement with the "real world" and assess the significance of the known limitations to the intended specific use at hand. It must represent how the model is intended to be used (by the decision maker) in the decision process, and it must assess the degree of risk in using the model in that way. The validation process is only the first step in such an assessment, and as such may miss the special features of the specific use. As a result, additional sensitivity analyses may be required in order to come to any conclusion before the official determination that a model is acceptable for a specific purpose.

Certain features from the definition of accreditation should be stressed:

It is official. The accreditation is performed by the decision-making official. This flows from the responsibility of the decision maker, which cannot be transferred to a computer, a computer model, or another individual. The decision maker must, *ipso facto*, believe in the tools used to provide the information to make the decision.

The operations researcher's task is to define clearly why to believe and how much to believe in this particular situation. The researcher has a right to know what is important to the decision maker. The transfer of belief requires a relationship between the researcher and the decision maker that is essential to the proper functioning of operations research. During the filtering that goes on as a report goes from the researcher to the decision maker, the critical assumptions and caveats too often get lost.

It is a determination. A decision has to be made, therefore accreditation is more than a process. How to make that determination should emerge from a dialogue between the operations researcher and the decision maker.

It includes a definition of what is acceptable. There must be criteria for "good enough" on which the researcher and the decision maker agree. The notion of "good enough" can be the subject of an analysis. It can be based on the consequences — the risk assessment for wrong decisions, or the cost of buying more confidence.

Accredit with respect to a specific purpose. Assessing the risks, costs, and consequences requires that the specific application be known. When the researcher does not know the use to which the research is put, the results of the research could easily be misapplied.

Accreditation, then, is an official decision that a model appears suitable to study a specific problem or issue. Accreditation goes beyond validation in that it makes a judgment taking into account the lack of "complete validation." During the accreditation process (which precedes the use of the model), there is clearly no decision to accept the model's results.

5.2 ACCREDITATION: THE NUTS AND BOLTS

As Quade and Carter note:

> . . .the argument for paying considerable attention to procedures for winning acceptance from the client and the staff as opposed to sole dependence on the logic of the analysis for that purpose is that if the findings of a policy analysis fail to influence the relevant decision makers, then that analysis, as a piece of policy oriented research, did not accomplish its purpose, no matter how good it might seem in the abstract or to other analysts.
>
> Immediate acceptance of all aspects of an analysis, however, is rarely to be expected; acceptance of ideas takes time. To be listened to and carefully considered is a practical goal, even though not a completely satisfactory one.[2]

To gain the acceptance we call accreditation, we recommend that the researcher start early, have a plan, have a team, have a methodology, and, in the end, have a document.

Accreditation should follow careful verification and validation. When can the operations researcher go to the decision maker expecting to get accreditation in writing? When can model accreditation reasonably be expected to be attained? Only after the results are in, when you have a chance to see what surprises may be in store. It would be irresponsible to accredit a model before seeing that the results make even the vaguest sense, or learning what aspect of the model drives the particular results that are supposedly relevant to the decision at hand. Accreditation should not be constrained to preset criteria, although criteria can be established as part of a specific accreditation effort. Does this mean that accreditation is of the model results? No. The results could turn out to be useless for the specific decision, even when those results are valid (known to be accurate to within a specified tolerance). This would be the case if the results were driven by what turned out to be a false assumption, or if the accuracy were not enough for the decision at hand.

The accreditation effort has to be tailored to the model application. For example, in the evaluation of a new weapon system, the accreditation effort would have to look closely at how new technologies are treated in the model, the impact on the model's decision rules given new tactics that are feasible with the new system, and changes in the existing modeled interactions between new and existing systems. What must be communicated to the decision maker are the strengths and weaknesses of the application of the model to a specific study; its limitations with regard to that application, and how those limitations affect the decision maker's risk of making a bad decision.

5.2.1 Start Early

Identify the specific purpose for which the model will be used. What is the question/decision? Keep in mind the story of the seven-year old who asked his mother "Where did I come from?" The mother had expected at least a few more years before addressing such questions, but, wanting to encourage openness and not stifle trust, she bravely explained all she could. After considerable time, the young boy interrupted, "Yes, but Bobby says he's from Spokane..." It is hard to give a meaningful answer if you don't know what the question really is. The most direct route is to find out from the one who must make the decision.

Identify the accrediting authority. From that person the researcher can find what are the variables of interest, what will be accredited, and what role the researcher's efforts will have in the decision (the latter is particularly important.) With this knowledge the researcher can focus: focus the model building, focus the verification and validation efforts, focus the sensitivity analyses, and focus the research on the real question and requirement for meaningful information.

5.2.2 Have a Process and a Plan

The plan should identify the issues and scope of the effort. The researcher should insist that the plan be reviewed and approved in the same chain that the accreditation will follow. The plan becomes a dry run of the process that will be followed after the results are in and accreditation is sought. This is a way of establishing contact between the researcher and the decision maker. There are, in many areas of model application, places where an accreditation plan can be described or referenced. For models to be used in support of operational test and evaluation, the COEA Guidance Memo and the TEMP are two appropriate places to seek agreement on accreditation issues.

The accreditation plan should include at least tentative criteria from the accrediting authority.

5.2.3 Have a Team

The researcher's efforts should make it easier for the decision maker (and the decision maker's advisors) to accept the results with knowledge of the strengths and weaknesses of the model.

One of the things that the researcher can do is to ask for a team to review the model. Strong teams have certain things in common. They are made up of experts. To ensure that they don't have a narrow slant on the subject, they are interdisciplinary. While the use of computer modeling may be new, the problem of building confidence is not. Aristotle suggests that to create confidence requires that the speaker appear to possess practical intelligence, moral excellence and good will. Interdisciplinary teams suggest "practical intelligence." Further, they don't have an axe to grind, which in bureaucratic parlance is often referred to as "independence." (They have "good will.")

Have someone from the accreditor's office on the team. Or establish a dialogue. Make progress reports...The team may be necessary because the model can have several users (decision makers) at the same time.

5.2.4 Have an Approach or Method

A model is a simplification of reality. It is used to represent some portion of reality. The first challenge that the researcher should address is whether the model is helping the decision maker to understand better. (The worst thing that could happen is for the researcher to take the view that the decision maker cannot or will not take the time to understand the results, how they were obtained, and why they are believable.) The purpose of the model is to improve understanding. The "answer" may be secondary. When this is the case, it is most important to explain the qualitative nature of the model output, and not to allow the decision maker to put undue confidence in the numerical outcome. It was the appreciation of this aspect that lead the Defense Science Board to warn, "Do not use models and simulations to prove things."

5.3 COMMUNICATION

The operations researcher has the following responsibility in communicating the results of the model: he must bring to the surface (for the decision maker or accreditor to see) those aspects of the model's employment and use that are most important in obtaining the results presented. The fact of the matter is that many complex models obscure seeing these "drivers." Exposing them is sometimes a difficult task. Without exposing them, the operations researcher's task is not complete. Good analysis requires such an effort. In the ideal case the results of the model can be derived on the back of an envelope (meaning that the result can be derived and explained to some approximation without the computer black box). If the researcher can, with the use of such an envelope, help the decision maker to understand what is "driving" the results, the work of transferring confidence to the model is very far along. The decision maker in general will have little trouble letting the computer generate the "next significant figure." Actually, the process of abstracting from the model the truly salient features, the driving factors, and the critical assumptions is what operations research should be about. If the real driver is the input that system x has an acquisition range twice system y, that should be explicit, not hidden in the million lines of code. Exposing why the model gives the result is critical to the decision to use or avoid use of the model.

In making the case for computer literacy, John G. Kemeny has noted:'

> Unfortunately, most decision makers in government and industry today are computer-illiterates. Although computer systems are in place in most large organizations, they perform mostly routine book-keeping functions and are used little, if at all, in decision-making. High-level executives, too embarrassed to expose their ignorance of computers by asking questions of the computer center, often leave important corporate decisions by default to computer programmers, who must fill in the gaps in the vague, general instructions they receive from top management.

The operations researcher should avoid "filling in the gaps" without educating the decision maker on how good the fill is.

Accreditation is a good place to start, as has been noted. There is no single method. There are some models of accreditation, but none is accredited. Each model has strengths and weaknesses which will be discussed. The accreditation plan should develop from the interaction of the researcher and the decision maker.

5.3.1 The Legal Analogy

Often models are used as part of a contentious, adversarial process. The model is used as part of the evidence for or against some particular point. This is generally unfortunate because, usually, the researcher who knows the work is generally not present to explain what was actually done with the model and what are the appropriate conclusions from the model runs.

If the expectation of the researcher is that such will be the case, one possibility is to encourage the debate in its proper forum, namely the research community. In such a case, for example, a red team could be formed to find and document the weaknesses of the model and its use. This would be submitted with the results of the run as part of the accreditation.

As with our legal system, we make the best case for and against the particular modeling or simulation application and pass judgment on whether or not to use it. Key to the legal analog are the following:

- present both sides fairly.
- have some ground rules for what is relevant, e.g., the MORS SIMVAL areas for consideration.
- do not suppress evidence; bring diverse views to the forum, expert witnesses, proponents, adversaries (the formation of Red Teams), whatever makes sense.

5.3.2 The Moral Analogy

While the legal analog may be necessary in particularly contentious cases, it does imply that there are two adversaries with positions to defend. So long as the debate focuses on the applicability of the model to the problem at hand, the debate can be helpful. Once the transition is made to the results, the analyst should be careful. "But of all our sins, the one that will finally hurt the profession the worst is the blurring of 'analysis' on the one hand and 'position-taking' on the other."[4] Protecting the profession is a noble goal; the analyst should not be a hired gun: "Have Model Will Travel." Self interest should also have a role. If the profession becomes litigious beyond reasonable limits, we will begin to share other attributes with lawyers. Decision-makers may come to feel as Shakespeare's Cade felt in Henry VI, Part II, "The first thing we do, let's kill all the lawyers".

Establishing trust, rather than inviting confrontation is not a new problem. Morse and Kimball note:

> The reaching of a working understanding on "terms of reference" between the operations research worker and the administrative head to whom he is assigned is one of the most important organizational problems encountered in entering a new field of operations research. Scientist and admin-

istrator perform different functions and often must take opposite points of view. The scientist must always be skeptical, and is often impatient at arbitrary decisions; the administrator must eventually make arbitrary decisions, and is often impatient at skepticism. *It takes a great deal of understanding and mutual trust for the two to work closely enough together to realize to the fullest the immense potentialities of the partnership.* (*Italics* added.)[5]

5.3.3 Test-Model-Test-Model

"The basic cycle of the scientific method may be divided into three steps: induction, deduction, and verification. ...*Induction* is the step which carries the scientist from factual observation to the formation of theories. Once the theory is formulated precisely, the tools of logic and mathematics are available to *deduce* consequences from it. Once a number of interesting consequences have been deduced from the theory, they must be put to the test of experimental *verification*..."[6]

Test results can be used to either further accredit or discredit a model (and by implication its results) by uncovering interactions and factors not foreseen in the modeling effort. One of the greatest dangers in model building is to ignore those factors that are difficult to model or for which there is scant data on which to base a model. Comparison with field results of any kind often provides the rude awakening that allows greater objectivity as to the limits of the model. There can be a certain tension between modelers and testers that can be of benefit to both. There is an old adage that, "No one believes a model, except the person who wrote it, and everyone believes a test result except the person who ran it." The modeler will look at the test result as one realization of possible outcomes and discount any discrepancy. This should not be encouraged. The model should produce estimates of variability also. The model should produce estimates of both the expected outcome, and the variability about the expected value. Any discrepancy should be examined to see if it is due to factors neglected in the model that contribute to variability, or to a sample size that was too small. This said, it should be clear why it is wise for the tester to choose the sample size with a knowledge of what the model predicts for the variability. Models that cannot be disproved based on test results are of a utility similar to tests in which it is impossible to fail: neither is worth considering.

5.3.4 Back-of-the-Envelope Believability

The purpose of a model is not to duplicate reality, the purpose is to increase our understanding of certain factors that are important in a problem. When a model becomes too complicated to explain, and the origin of results is obscured inside a black box — then the model has not increased our understanding, it has obscured our ignorance. (Often this is evident in an exchange that goes: "Why did it turn out that way?" — "That's just the way it turns out!") The model should allow traceability of cause and effect between the variables of interest and the outcomes.

One technique or discipline that can be used is to strip the model down to essentials, find out what are the key assumptions that drive the model, and develop a simpler model that is easier to understand and explain. In fact, good research might begin with the simplest models that are back-of-the-envelope. More complicated models develop as more factors that are potentially relevant are treated more explicitly. These additional factors may change detailed numerical outputs but should not change overall trends or conclusions, provided the original model was good. Whichever way the chronology of the research occurs, the result can be a hierarchy of explanations that go into greater and greater detail, until the results are understood to a point where the decision maker can understand why and how the results come about.

Often the operations researcher is not the developer of the model, but is expected to employ a model and use it to get an answer. This is a very dangerous situation. The researcher must first determine, in the researcher's own mind, the acceptability of the model for the specific purpose. The technique described above works particularly well in such a situation.

5.3.5 Risk Assessment as Part of Accreditation

The level of effort applied to accreditation is driven by the perceived importance of the use of the model. Accreditation should demand that an analysis of potential "unknowns" be done and documented.

One standard technique that the operations researcher has is Decision Theory. In such an approach the analyst will evaluate the consequences of making a mistake by using the model. The risk will depend on the model and the decision to be made. It may depend on the phase of the program.

Table 5-1. Example of Risk Variance in Use of Models in the Acquisition Process

Phase	Area of Impact	Impact
Mission area analysis	Further studies	
COEA	Choice of alternatives	Keep one or more alternatives
System Design	Engineering Analysis	Choose to run tests
Test planning	Sample size	Cost of test
Test execution	Shape battle	Test realism
Evaluation	Pk analysis	Exit criteria
Milestone III	Procurement	Fielding

Risk assessment could work as follows:

- First, together with the decision maker determine the purpose to which the model will be put, and the decision actually to be made.

- Then develop a model (a meta-model?) of how the computer model will be used in the decision process. (For example, it might be to confirm that no previously used models contradict what the decision maker wants to do, or it might be to generate a single parameter estimate that is a go-no go criterion.)

- Assess the risk (expected loss) in the use of the computer model by determining (1) the probability of a wrong answer from the model (for example due to the error bounds of the inputs), (2) the effect of a wrong answer on the decision, and (3) the ultimate consequences a wrong answer might have on the program. Some decisions, like investment decisions, cannot be avoided, but the program can be corrected if later events do not confirm the expected outcome. Other decisions might be uncheckable at all, or only after huge resources have been wasted. In examining this aspect of the question, the adequacy of the decision makers plan for the program needs to be understood. In fact, the plan for the program should have been constructed with internal checks so that wrong decisions can be found and corrected without significant loss. In assessing the risk at a particular decision point, the analyst may be able to use the planner's analysis if it is available. During defense program execution, there are distinct phases during which the consequences of a wrong decision are very different (see table).

For example, the accreditation process may determine that the model does better in determining relative differences than absolute values. The model suggests that one alternative is preferred. If the uncertainty in the absolute results of the model is so great that neither alternative considered may satisfy the need, the decision maker should know that the mission need may go unfulfilled with some probability if a single preferred alternative is chosen, and a different probability if two alternatives are kept under development. The cost of keeping one or two alternatives under development must also be considered. The cost of an extra alternative through demonstration and validation may be small compared to incorrectly choosing a preferred alternative and not finding out until all the development money is spent. In other words, knowing how much to believe the results of the model may allow the decision maker to hedge the risks. It may encourage exploration of alternatives with less risk. It may encourage a program modification to gather the kind of data so that a better model could be developed. It may encourage the decision maker to insert test points into the program in order to gain more confidence that the program will eventually pay off.

As a second example consider that using a model to help plan tests can be very useful, and not incur great risk. For example, a model can be used to estimate the variability of test outcomes in order to help determine an appropriate sample size. If the estimated variability is wrong, the confidence level of the test may be changed, and if the variability is very wrong, the test may have to be extended. But the test results will still be available. The knowledge that there were sources of variability that were not accounted for in the model will probably stimulate an improvement to the model, and warn the decision maker about the model.

5.3.6 Documentation

The emphasis on accreditation today means that the operations researcher should document the evidence, the review process, and the thought process. The table below suggests what the documentation should include.

Table 5-2. Considerations for Accreditation Documentation

Evidence	What is the evidence that the model is appropriate for the problem at hand? Are the input data to the model relevant? What is the data base on which the model is built? Is this data base relevant?
Criteria	What criteria were used to decide on the appropriateness of the model?
Process	Who is the decision maker or accreditor?
Decision	What [illegible] the decision maker need in order to understand [illegible] purpose of this accreditation?
Caveats	What warnings need to be clearly stated?

5.4 SUMMARY

Reduced to its essentials the VV&A problem for the operations researcher is the following:

- All models are wrong (at some level and in some way).
- Validation determines how the model is wrong and when (i.e., it determines the limits and errors in the model.)
- Accreditation is the determination that the decision to be made is not sensitive to those errors and limitations.

The approach to accreditation outlined here is not institutionalized in directives or plans. What is suggested is that, as a matter of professional practice, an operations researcher with a modeling problem should actively seek to install an accreditation framework within the project. This will allow the development of a focused and mutually beneficial interaction between the researcher and the decision maker. The role of any model is to increase understanding and facilitate communications. The model is a tool of, not a substitute for, good judgment.

Endnotes

1 E. S. Quade, *Analysis for Public Decisions*, Third Edition Revised by Grace M. Carter, 1989, North-Holland, New York, p. 169.

2 Ibid., p. 332.

3 John G. Kemeny, "The Case for Computer Literacy," *Daedalus*, Spring 1983, p. 218.

4 Glenn Kent, "The Role of Analysis in Decision Making," Keynote speech before the 24th meeting of the Military Operations Research Society, 1969.

5 *Methods of Operations Research*, Philip M. Morse and George E. Kimball 1946, Operations Evaluation Group, Office of the Chief of Naval Operations, Washington, D.C., pp. 8-9.

6 *Mathematical Models in the Social Sciences*, John G. Kemeny and J. Laurie Snell 1962, Ginn and Co., Boston, p.3.

CHAPTER VI - A FRAMEWORK FOR VERIFICATION, VALIDATION, AND ACCREDITATION[1]

Paul K. Davis

6.0. PREFACE

This study was developed for the Defense Modeling and Simulation Office (DMSO), which is under the Director, Defense Research and Engineering. The study reflects discussions of the DMSO's Applications and Methodology Working Group, which I chaired during this work. The study also draws upon discussions at two special meetings on verification, validation, and accreditation (VV&A) sponsored by the Military Operations Research Society (MORS) on October 15-18, 1990 and March 31-April 2, 1992. Nonetheless, the material presented here is my responsibility and I make no claims about consensus in the community. VV&A is a difficult subject on which there is a broad range of opinions and practices (e.g., VV&A of software used in space probes is different from VV&A of military simulations used for analysis). At the same time, it appears that a considerable convergence of view is taking place and I hope that this study will accelerate that process. Comments and suggestions are therefore especially welcome. They can be sent by electronic mail to Paul_Davis@rand.org through Inter Net.

Work on this effort was accomplished in the Applied Science and Technology program of RAND's National Defense Research Institute, a federally funded research and development center sponsored by the Office of the Secretary of Defense and the Joint Staff.

6.1. INTRODUCTION

6.1.1 Objectives

Verification, validation, and accreditation (VV&A) is a complex subject that has troubled developers and users of models for many years. Each generation of modelers and analysts must think it through, because understanding the issues is important to professionalism. Consumers of analyses exploiting models must also understand the subject or they will have difficulty judging the quality of products. Further, they may either be insufficiently demanding or supportive of VV&A efforts on one extreme, or unreasonable on the other—requiring a degree of validation that is impossible even in principle. Managers of analysis organizations should understand VV&A so that they can put into place appropriate procedures, standards, and incentives. This may be called a VV&A "regime" to emphasize that VV&A is not a one-time event, but rather an ongoing but episodic *organizational* activity that should be understood and considered important by all participants.

What, then, might a VV&A regime look like if one saw it? What advice should be given to a new manager who is ready and willing to institute reforms to establish sound VV&A policies and procedures? This study is an effort to sketch the essential features of an answer. Its principal objective is to provide guidance that would be useful to such a manager in government, industry, or the academic world. Auxiliary objectives include discussing the special VV&A problems associated with knowledge-based models and recommending new attitudes about model development and VV&A that reflect implications of modern technology.

6.1.2 Background

There is a considerable literature on VV&A for military models, much of it severely critical of model developers and their government sponsors for there not having been enough VV&A in the past.[2] There is no definitive source on what VV&A is or should be, but someone new to the field might well consult Thomas (1983), other chapters of Hughes (1989), Gass (1982), Sargent (1987), and Martin Marrietta (1990).[3] The first of these has a philosophical slant and addresses some of the profound difficulties in even contemplating model evaluation. The latter, which draws on the work of Gass, Sargent, and others, describes an approach that has been used in large-scale efforts having to pass rather stringent DoD criteria. Another good introduction to validation issues is Miser and Quade (1988). Finally, those concerned with VV&A will surely want to examine guideline documents emerging from sponsoring organizations, as well as regulatory documents such as U.S. Army (1992) (especially Chapter 6 on VV&A) and DoD-MIL-STD 2167, which describes software standards.

In this study I present some definitions (Section 2) along with discussion of what the definitions mean and why they are not simpler. My definitions of validation and accreditation extend the more usual ones in important ways. Section 2 then presents a taxonomy of VV&A methods, focusing primarily on validation. Section 3 describes VV&A as a dynamic *process* that should conduct evaluations for both broad classes of model application and for specific studies having detailed analytic plans. Section 4 then pulls things together and recommends an approach for the use of practitioners, managers, and consumers of model-based analysis.

6.2. DEFINITIONS AND CONCEPTS

6.2.1 Models and Programs

"Models" are representations of certain aspects of reality (e.g., of certain aspects of particular systems). They come in many forms, including the physical scale models used by architects, analytical models expressed in paper-and-pencil equations, and computer models (see also the overview chapter of Hughes, 1989). This study focuses on computerized models, primarily "simulation models," which attempt to describe how a system changes (behaves) over time.[4] I am also concerned here with models having phenomenological content relating causes and effects rather than, say, regression "models" or optimizing algorithms that some might call models.

Although the terms "model," "simulation," and "program" are often used interchangeably, here and elsewhere, it is sometimes important to make distinctions, especially between the model (or what some call the conceptual model) and the program (or computer code), which implements the model. Annex A elaborates on this and argues, reluctantly and in contradiction with the advice given by most scholars, that it is becoming increasingly difficult—and decreasingly appropriate—to separate the processes of designing and evaluating models on the one hand, and designing, building, and evaluating program implementations on the other. Technological change demands a new approach here.

6.2.2 Models, Data, and Knowledge Bases

Throughout this study "model" means the union of a "bare model" (also referred to as "the model itself") and its "data base." Thus, $Y(t) = Y(0) - 1/2\ g\ t^2$ is a bare model, while $\{g = 32\ ft/sec^2;\ Y(0) = 10{,}000\ ft\}$ is a data base. In some instances, the data is a "knowledge base" in the form of rules and algorithms.

In the past, bare models were conceptually distinct from data in most cases. The bare models defined structure and algorithms, while the data base provided values (e.g., for the gravitational constant or the number of tanks in a division). Modern practice, however, has muddied the distinctions. In many models, much of the substantive content is defined in the data base because with most computer models it is easier and faster to change data than the program itself and developers have sought to provide users as much flexibility as possible.[5] As a result, *the VV&A process must consider both bare models and data bases.* [6]

Quite often, bare models and data bases need to be reviewed together, in the context of an application; in other cases (i.e., with different model designs), they can to greater or lesser degree be reviewed separately. For example, one can conduct VV&A on an order-of-battle data base without knowing precisely how that data base will be used. Similarly, one can conduct VV&A on an algorithm without knowing the precise context in which it will be used.

6.2.3 Verification

> Verification is the process of determining that a model implementation (i.e., a program) accurately represents the developer's conceptual description and specifications.

This is the definition commonly accepted in the military modeling community. There continues, however, to be some confusion and disagreement about precisely what is and is not covered under verification, and about what taxonomy describes verification activities. I consider verification to consist of two basic parts.

- *Logical and mathematical verification* ensures that the basic algorithms and rules are as intended by the designer and do not include logical or mathematical errors (e.g., divisions by zero, incompletely specified logic, or nonsense results when certain variables take extreme or unusual values). Although verification is nominally concerned with implementation rather than correctness of design, it is common for verification activities to uncover design errors along the way (e.g., to detect an implicit and unreasonable assumption about independence of events). Thus, verification activities should begin with documentation and will often accomplish some validation functions.

- *Program verification* (or *code verification)* ensures that these representations have been correctly implemented in the computer program. Program verification is concerned in part with simple matters such as discovering and correcting typographical errors, errors in the units in which physical quantities are described, and errors of definition (e.g., a model designer might have intended that a force ratio apply only to forces on the forward line of troops (FLOT), but the programmer might have defined it to apply to groupings that include corps-level reserves). It is also concerned with more complex issues such as the appropriateness of numerical integration techniques,[7] covering all the logical cases (including cases that the designer might consider unlikely or unphysical), and eliminating bugs that would cause the program to "crash" in some circumstances. Many such bugs involve intricacies of the particular computer hardware, operating system, and interface software.

Verification is a matter of degree for complex models, because it is impossible in practice to test the model over the entire range of variable values and because it is often not feasible with available resources to do a line-by-line code check. Thus, a model may be well verified within a particular "scenario space," but not well verified otherwise.[8] In principle, one might think of using sampling techniques to verify code to some level of confidence, but I am personally unaware of any rigorous efforts to do so in the realm of combat modeling.

Verification of *data* (especially classical types of data such as physical constants or orders of battle rather than, say, data defining elements of model structure or exponents in algorithms) should often be distinguished from verification of the bare model, because different techniques are involved and data bases change frequently.[9] There are at least two aspects of data verification. The first aspect involves ensuring that source data are converted properly to model input data and are consistent with the model concept and logical design (e.g., that data supposed to represent conditional probabilities of kill given a hit do indeed represent those rather than, say, kill probabilities per *shot*). It should also include spot checks to confirm that data were, in fact, extracted from the stated source and that it represents the latest available from that source. If data is not provided with the model, then verification should include establishing that the required user inputs are readily available.

A different aspect of data verification applies within the context of a study if the data base has already been installed. Here one seeks to establish whether the data base represents correctly the assumptions intended for the analysis. For example, if an analyst states that he wants to use a particular official data base for orders of battle, data-base verification would include checking that the desired data base was the starting point for the installed data base, but it would also check to see if appropriate corrections had been made—corrections that the analyst would surely want if merely he knew to ask for them. These would include providing realistic data values where the original data base had zeros, blanks, or values annotated as "purely nominal." Official data bases are often riddled with holes and errors. Managers of analysis and recipients of analysis are often unaware of how serious these holes and errors are, or of how much the analysis depends on the cleaning-up process, which often requires substantive work and numerous subjective judgments (which unavoidably mixes verification and validation activities).[10]

6.2.4. Validation

> Validation is the process of determining: (a) the manner in which and degree to which a model (and its data) is an accurate representation of the real world from the perspective of the intended uses of the model and (b) the subjective confidence that should be placed on this assessment.

This definition extends the more conventional definition.[11] The extension calls attention to two considerations. First, there are different meanings to "accurate representation." Second, the validation process should address the issue of confidence (not in the sense of "statistical confidence," but in the larger sense having to do with how much one would bet on the correctness of the model's predictions given residual uncertainties). While one could consider both considerations to be implicit in the more usual definition, it seems to me evident from experience that they will be underappreciated unless made explicit.

Types of Validity

To elaborate on the definition given above for "validation," I use the phrase "manner in which" because a model can be "valid" in several distinct ways. It may have (a) descriptive validity, (b) structural validity, and/or (c) predictive validity (see also Zeigler, 1984).[12]

Descriptive validity means here that the model is able to *explain* phenomena or organize information meaningfully in one way or another. For example, a descriptive model might be able to say, "Well, the reason this happened is that A collided with B, which happened because A had lost its radar and therefore failed to see B in the cloud bank." All of this might be a sound and nontrivial reconstruction of events. Note that the model used for such a reconstruction might not have been able to predict the events ahead of time, especially if the key causative events were stochastic or some key inputs such as precise speed histories were unknown. What constitutes a "good" description or explanation depends on context and taste.

Structural validity means that the model has the appropriate entities (objects), attributes (variables), and processes so that it corresponds in that sense to the real world (verisimilitude), at least as viewed at a particular level of resolution.[13] One may also require, for structural validity, that the principal algorithms are at least roughly appropriate, although not necessarily accurate (e.g., whether a process describes exponential or linear growth may be regarded as a structural issue).

Predictive validity means that a model (including available or potentially available data) can predict desired features of system behavior, at least for particular domains of the initial conditions and durations of time, to within some known level of accuracy and precision. A conditionally predictive model explicitly identifies alternative behaviors and the conditions that would cause them (e.g., "*If* the weather tomorrow remains clear, *then* the air operation should go well and... ."

These types of validity can be considered more or less orthogonal attributes of a model. As suggested in **Table VI-1**, one can have models with every combination of type validity. This is significant to VV&A, because the criteria one applies depends strongly on the type of validity sought.

Table VI-1 Models With Different Combinations of Validity Type

Case	Descriptive Validity	Structural Validity	Predictive Validity	Example
1	Yes	Yes	Yes	Well-tested weapons-performance models.
2	Yes	Yes	No	Good-theater level models (which may, however, be conditionally predictive for some features of a campaign, at least in certain domains such as when one side has overwhelming force).
3	Yes	No	No	Historically based statistical models correlating different measures of outcome (e.g., movement rate and ratio of loss rates; one might say "Because the ratio of loss rates was low, the advance rate was fast."
4	Yes	No	Yes	Some highly aggregated models that reflect doctrine and experience (e.g., march times for unopposed moves)
5	No	Yes	Yes	Incomprehensible but reliable black-box models with high resolution in entities and processes (e.g., poorly coded models with little documentation or explanation capability).
6	No	Yes	No	Models with high resolution in entities and processes, but poor algorithms (e.g., weapon-on-weapon attrition calculations assuming perfect tactical command and control).
7	No	No	Yes	Rules-of-thumb models or statistical models that work for no clear reason and do not represent system structure (e.g., a regression model predicting the next week's weather as a function of today's weather and the month of year).
8	No	No	No	Bad models.

To illustrate a few points in **Table VI-1**, consider first that a model can be excellent, even definitive, for explaining phenomena *after the fact*, and yet be useless for prediction (e.g., Case 2), at least in the usual sense. This happens if the model depends on the values of variables that are unknown until after the fact (e.g., the fighting quality of the other sides' forces). This situation occurs commonly with military models, since we do not know the detailed initial conditions for future military operations. Nor do we know the various decisions that will be made in the course of the operations. After the fact, these decisions and other previously unknowable variables may be unambiguous and objective data (e.g., as reflected in operations orders and reports on what the weather was). If the model then explains the phenomena well in retrospect (sensibly as well as accurately), the model is descriptive. [14]

As a second example, structural validity does *not* imply that the attribute values are correct or that the algorithms constituting the model processes are precise. A model of combat might be structurally valid while treating attrition quite approximately: it would have an attrition process, but the process would be inaccurate (Case 6).

The most subtle example here is probably that predictive validity does not imply descriptive validity, in our sense. One can have an empirically-based model, perhaps in statistical form, which is remarkably predictive, but which says little or nothing about the cause-effect relationships at the levels of physical entities and processes (Case 7). It is often difficult to know when such models will fail, but they are useful nonetheless.[15] [16]

Again, then, the point here is that evaluation of models should vary with type. It is silly to denigrate a good descriptive model that is structurally valid, merely because it is not a prediction machine (given the data known ahead of time). This is nontrivial, because many critics of military modeling are guilty of precisely this error. Those who argue that attrition estimates for the Desert Storm operation were off by an order of magnitude overlook the fact that many analysts were explicit about their estimates being upper bounds and about the potential for much lower attrition if the Iraqis proved ineffective by virtue of poor morale, training, leadership and so on.

Issues of Degree and Confidence

The words "degree" and "confidence" appear in my definition of "validity," because models are seldom perfectly valid in any of the dimensions (description, structure, or prediction). They vary in their accuracy and precision. Also, there are several dimensions of confidence, since:

- The model or its data may be known to be highly uncertain (e.g., in functional form or in data values).

- The model and its data may represent a best-estimate consensus of experts, but may nonetheless be fundamentally wrong (e.g., Ptolemaic astronomy). One dimension of confidence, then, relates to assessing the likelihood of the bare model or its data having serious flaws that have not yet been thought of or taken seriously.

- A model may be deterministic, while the relevant world may be stochastic.[17] In this case, confidence in the model's predictiveness depends on the underlying probability distributions. If the distribution function is strongly weighted around a central point, then a deterministic model may be reasonable; if the function is bimodal, then such a deterministic model may be downright misleading.

For all of these reasons the process of validation should include reaching explicit, albeit often subjective, judgments about the confidence one places on the model. These can be aided by sensitivity analyses coupled with analysis assessing how much one truly knows about the more critical variables in the context of a shooting war.

Some examples may be useful here to illustrate how central the issue of confidence really is in the use of military models. Consider the following hypothetical statements about models being made by analysts to general officers in the context of a real war or preparations for such a war:

> The strategic-mobility model itself is solid, for aggregate predictions, but predictions depend on planning factors and decisions. We should plan for buildup rates +/- 30% around baseline data. Also, we should recognize that the CINC may make significant changes in the Time-Phased Force Deployment List (TPFDL), so we must anticipate the kinds of changes he would most likely seek and consider their consequences on predicted buildup rate.

> Because of uncertainties, including random factors and intrabattle decisions, we have no confidence in predicting winner or loser (or low casualties)—unless we can stack the deck by going for a 6:1 local force ratio after bombing. Then we would be confident.

> Results will depend on surprise and speed. That's beyond our model's ability to predict well. The model is descriptive after the fact, but that doesn't tell us what we need to know now. We can instead tell you, as a commander, how quickly we think you need to maneuver for success, based on intelligence estimates on the enemy's reaction times and maneuver speeds as judged from doctrine and exercises over the last few years. Whether you can do that is difficult for us to judge.

> The ECM-ECCM model is very accurate for aircraft flying against the SA-99 as we know it, but the enemy may have changed subsystems, in which case noise jamming would be unchanged but false-target generation might not work at all. We simply don't know whether he has changed systems.

All of these statements could be made quantitative to avoid ambiguity, but my own recommendation is to use the language of odds in a context that downplays confidence and reminds everyone of the stakes (e.g., mens' lives) rather than using the language and tone of statistical precision. As an example:

> If we have characterized the SA-99 correctly, as we *think* we have, our ECM should be less than 1% (between about 0.5% and 1 %). If the enemy has changed subsystems and can defeat our false-target generation (this is highly subjective, but I'd say that's a 1-in-4 situation), then our rough calculations suggest our losses will be about 1-2% per sortie until we can destroy the SAMs. Even in the bad case, we estimate that losses won't be worse than 3-4 % per sortie because they have a limited number of SAMs. That loss rate might last up to three or four days, but we're very confident we will destroy the SAMs in no more than that time.

<u>Data Validation</u>

In most of this study I treat data validation as part of validation generally. It is worth mentioning some unique features of data validation, however. These relate primarily to the types of data one uses to introduce facts, official estimates, and other numbers rather than, say, the types of data one may use to define aspects of the model (e.g., spatial resolution or exponents in equations). In this activity one typically reviews the data sources and how they were collected to compare model input data to real-world or best-estimate values. This may involve assessing the credibility of data sources and comparing alternative data bases. In reviewing operational data, one must consider exercise artificialities such as safety-related constraints and geography. Data validation is often quite troublesome. Intelligence estimates, for example, may vary widely with little rationale given and estimates of system effectiveness for U.S. weapons are often extrapolations from small data samples collected under artificial conditions.

6.2.5. ACCREDITATION

> Accreditation (often used synonymously with certification) is an official determination that a model is acceptable for a specific purpose (e.g., to a class of applications or to a particular analysis or exercise,)

<u>Accreditation By Class of Application Vs Specific Application</u>

Except for the parenthetical phrases, this is a commonly accepted definition (e.g., Williams and Sikora, 1991 and U.S. Army, 1992). It says that accreditation is a *decision* (not just a process) to the effect that a given level and character of verification and validation are sufficient to justify using a model in a particular application.[18]

Problems arise not with the definition, but with what organizations charged with model VV&A sometimes try to do. It would be convenient for such organizations if models could be definitively accredited for broad *classes* of applications, but even within a given class of applications (e.g., weapon-system comparisons), a model will sometimes be adequate and sometimes not.

Which situation applies depends on details, including numerical details and the sensitivity of results to errors in model performance. Also, some models that might be thought inappropriate to a particular application can be used effectively if manipulated cleverly with the benefit of parametric variations informed by side calculations.[19] *It follows that class-level accreditation should be provisional only, and that accrediting authorities should be extremely cautious in claiming that models cannot or should not be used for applications within a given class.* Those long familiar with VV&A issues and organizational behavior are perhaps most concerned about this problem, because they see the potential for mischief when controversial studies use models. Another concern here stems from the observation that organizations sometimes insist that "accredited models" be used for studies even when those models are inappropriate compared to alternatives that have not yet been accredited, or even fully developed. Furthermore, many fear that the accreditation process will place too much of a premium on verisimilitude and too little emphasis on clarity, controllability, and efficiency.

A Crucial Issue in Sound Accreditation: Model Clarity

It is perhaps a symptom of the disconnect between analysts and those who build and sponsor models that discussions of VV&A seldom mention one of the most important considerations in evaluating a model: its *clarity*. One could argue that the definition of validation should be modified to include such considerations, but I have chosen in this study to argue that these considerations are very much in the province of those who oversee particular uses of models. They have an important stake in model clarity, because:

- They are responsible for results and their ability to review the work (or have it reviewed by independent experts) depends on their ability to comprehend the model and the cause-effect relationships dominating results.

- They are responsible for communicating results, which typically requires separating essentials from noise.

- They may want to be able to reproduce the work, which will be far easier if it has been conducted with a comprehensible model.

It follows, then, that accreditation should depend not only on the soundness of the model for the application at hand, but on the ease with which the model can be comprehended and the results of the model understood in terms of appropriate cause-effect relationships. *That is, model accreditation should depend not only upon model soundness for the application, but also upon: (a) comprehensibility of the model and (b) comprehensibility of model runs (through "explanation capabilities").* This facet of the problem has been greatly underappreciated in prior discussions of VV&A, even within the academic community and even by systems analysts, who certainly wax eloquent about the need for model simplicity in other contexts. I observe also that the importance of model clarity increases the importance of establishing a model's descriptive validity.[20]

6.3. A TAXONOMIC VIEW: THE CONSTITUENTS OF VV&A

6.3.1. Prefatory Distinctions

Given the above definitions, how

does one *accomplish* VV&A? Suppose one is attempting to establish a *VV&A regime* within an organization, a regime in which one routinely does virtuous evaluation before using models for analysis. How does one go about it?

It is useful first to make some distinctions:

- Components vs. system (or modules vs. integrated model)
- Bare models vs. data
- Evaluating "best estimate" functional forms and data values vs. evaluating ranges, distributions, and confidence
- Conducting "broad VV&A" with only a partial sense of the intended applications vs. conducting focused VV&A for a particular study[21]

VV&A applies to each half of each of these pairs. I emphasize this up front, rather than repeating it at every point of the following discussion.

6.3.2. A STRUCTURAL PERSPECTIVE: THE COMPONENTS OF VV&A

Figure VI-1 now provides a *structural*, or taxonomic, view of what constitutes VV&A. It elaborates on validation, because that aspect has been most controversial and confusing over the years. I use the phrase "generalized validation" or "evaluation" here, because my sense of validation is broader than that of some authors.

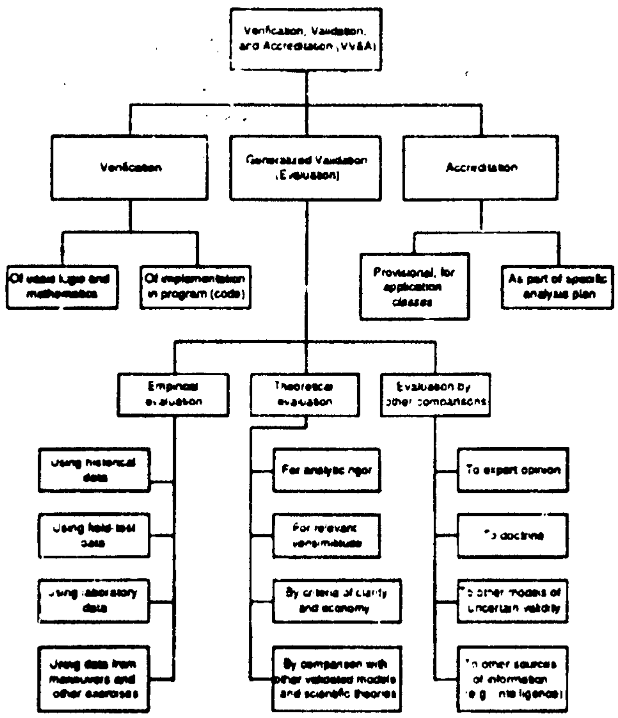

FIGURE VI-1. A Taxonomic View of VV&A

6.3.3. Verification Methods

Although this study does not emphasize verification methods (see Sargent, 1987 and Martin-Marrietta, 1990 for more discussion), the traditional methods include (a) walking through the design and code; (b) studying flow diagrams; (c) checking algorithms; and (4) using CASE tools. Significantly, modern software methods coupled with the development of expert systems to assist verification can greatly improve the quality of models and the efficiency of the verification process (e.g., by detecting errors when they are introduced). Many of the methods seem mundane when described, and may seem burdensome to those who must do the typing of code, but they are exceptionally powerful and have not yet been fully exploited. Examples with which I am personally familiar include:[22]

- Strong typing in computer languages, which detects a wide variety of typographical errors and ambiguities such as having different names for the same variable or different variables with the same name.

- Range constraints on variable values, which are entered (as data) at the time variables are declared and which allow the executing program to become aware of likely errors (as evidenced by variables taking on values outside the prescribed ranges) and to print error messages.

- Automatic testing for logical completeness in decision tables and equivalent sets of If-Then-Else loops.

- Well structured "explanation logs" at alternative levels of detail, which allow a reviewer quickly to scan not only final results but values of intermediate variables and the logical paths being taken in the simulation.

- Use of object-oriented design methods, which, when physically natural, provide improved modularity and better organized data structures that simplify verification.

These techniques[23] can be especially useful for verification of implementation in code, but can also be useful in highlighting spurious logic (e.g, in explanation logs).

6.3.4. Validation Methods

Validation as a Holistic Process

Most experienced modelers and analysts consider validation to be a holistic evaluative process that includes many different kinds of testing. Some of this may be classic empirical testing of the sort often associated with the scientific method. In practice, however, it is only rarely possible in policy analysis to conduct the controlled experiments necessary for such rigorous testing of the model as a whole. Where such experiments are feasible, they should be greatly valued, but we cannot conduct controlled wars or even perfectly controlled battles (nor can we conduct perfectly controlled social experiments on matters such as health care options).

We must settle for something a good deal less than idealized scientific rigor.[24] Nonetheless, there is ample opportunity for empirical work. As suggested by the empirical-evaluation column of **Figure VI-1,** some aspects of models can be tested or informed by comparisons with historical data, field-test data, or data from operational maneuvers and other exercises. This data is not usually as well controlled or as directly relevant as one might like, but it is very useful nonetheless.

Looking to the central column of **Figure VI-1**, other less empirical methods should be key players in generalized validation. The first is theoretical analysis (e.g., working through the substantive logic, checking relevant verisimilitude, considering the reasonableness of assumptions, applying criteria such as requiring falsifiability[25] and the use of Ockham's razor, and comparing assumptions and implications of the model with well established theories from physical science, engineering, and military science[26]). Theoretical analysis, then, goes well beyond what is suggested by the phrase "logical validation," which sometimes appears in discussion of VV&A (e.g., Williams and Sikora, 1991). Theoretical analysis often exploits special cases in which it is possible to compare the model in question with exact calculations based on rigorous or otherwise well established theories.[27] Sargent (1986, 1987) lists some of the various methods that can be used in this connection.

Looking to the rightmost column of **Figure VI-1**, there are a variety of other comparisons one can make to evaluate a model. These include comparisons with expert opinion, doctrine, and so on. Finally, **Figure VI-2** emphasizes that these evaluations all feed into an overall evaluation holistically. There is no natural order or ranking of evaluation methods, despite efforts to create one (e.g., as discussed ambivalently in Williams and Sikora, 1991, although subsequent MORS works has dropped the effort to impose an order). This is not entirely trivial, since false ideals cause trouble and the ideal of believing, for example, that data from maneuvers is the "best" and "most important" data to be used in validating a model will typically be wrong. Basically, model development and evaluation involves using many sources of information and tying it together however one can. It is not so orderly as some would have it.[28]

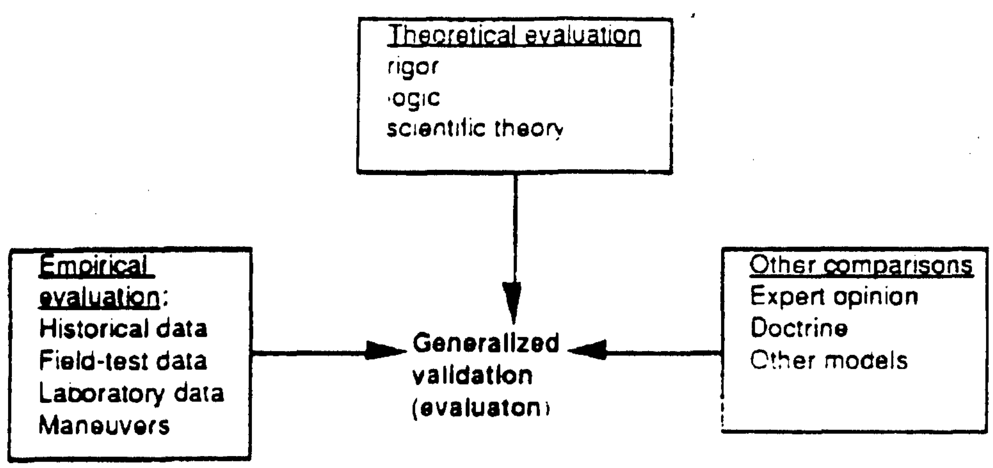

FIGURE VI-2. Validation as a Holistic Process, Not a Linear Process

A Perspective on Validation

It is sometimes useful to think about validation as an informal application of Bayesian reasoning under circumstances in which we can only estimate the probabilities. Our objective is to develop representations that are good enough "to bet on;" but we will seldom have a sure bet and we the refore want to have a sense of the odds for each of a number of very different kinds of wagers.[29] This validation process is unquestionably subjective, but not capriciously so. We consciously seek information that could falsify or reinforce our judgments and we attempt to face up to that information when we obtain it. When all is said and done, however, we must *do something*. That is, we must conduct the best analysis possible given the information, time, and resources available to us. Ultimately, *validation (and accreditation as well) is concerned with establishing that we are indeed doing the best we can*--or, at least, something that is "good enough." It cannot be separated completely from context.[30]

Issues of Breadth and Depth in Model Validation

A model's validity is one thing; the extent to which it has been validated is another (i.e., a good model may not yet be known to be good). A common problem for those overseeing the development and use of models is "How much validation is enough?" Another question is "How do we start?" **Figure VI-1** provides a checklist of methods, but most of them could become lifetime careers when dealing with complex models. It is therefore useful to make some further distinctions, which also have the effect of suggesting where to start.

Depth in Validation. As with most human endeavors, the value of validation activity is described by a curve of marginal returns—a curve that rises steeply and then begins to level off and move slowly toward an asymptote (which may correspond to considerable, and yet incomplete, confidence). For a variety of reasons, some of which could probably be explained theoretically, it seems to be the case that even a little validation can go a long way. It is for this reason that "face validity assessments" are so important in practice. These can be attempted in each and every validation-related box of **Figure V-1**. Some examples will probably convey the ideas. Once again I use the technique of plausible statements that might be made in characterizing a model's validity:

> *Using historical data.* The model is absurd. It took me all of 30 seconds to discover from the output graphics that it has field armies moving at an average speed of 150 km/day over the course of a successful ten-day campaign. Probably, some nitwit physicist built the post-breakthrough movement algorithms after thinking about how fast tanks can drive. Historically, opposed movement has been more like 20 km/day, although there have been special cases.[31]

> *Using field-test and exercise data.* The model is exceedingly optimistic about the effectiveness of TOW missiles (kills per shot and shots per battery per battle), probably because of using test-range data uncritically.

Results from the National Test Range and Desert Storm give a very different picture.

Using simulator data (a kind of laboratory data). The model for pilot acquisition rates in finding mobile targets is in fact more reliable than what the pilots are telling us anecdotally based on normal training practice. There have been some experiments in simulators that demonstrate pilots are much more conservative about declaring a target detection when they are concerned about friendly forces being in the region or about hitting civilian targets. In terms of the required signal-to-noise ratio, the difference is...

Testing for analytic and scientific rigor. I quit reading the documentation as soon as I discovered that the detection model assumes a uniform background over areas as big as middle-eastern countries. We know that the ability to track a target (not just detect it once) depends on being able to maintain a reasonable signal-to-noise ratio, and that background varies substantially over distances of tens of meters, even in the desert. I also note that the model ignores the effects of cueing and prior knowledge by using independent probabilities. We need a better acquisition model.

Looking for relevant verisimilitude. The model treats logistics quite crudely, at the level of tons per day of consumption, tons on hand (by sector), etc. However, it looks about right in aggregate: divisions in intense combat use about...tons per day, but intensity seems to drop pretty quickly, which is reasonable. The real problem is that there is no mechanism in the model for one side to affect the other side's supply capability. The model is structurally unsound in that respect. It doesn't even model support units and allow attacks on their trucks.

Evaluation for economy. The model may or may not be accurate if one knows all the input variables precisely, but it's going to be impossible to use well for systems analysis in realistic cases where we don't know those values in many cases. The model has so many tuning parameters it could fit anything after the fact, but I don't think it's worth much for our purposes.

Comparisons with familiar models. Well, it's a different model, of course, and there are scores of parameters that I didn't try to review in detail, but the model at least behaves reasonably in the sense that it gives the same picture of what would happen in the several baseline cases of the ...study as came out of the full-up war game at CINC headquarters.

All of these examples could have been the result of fairly casual checks of face validity by different experts. None involved detailed testing. In my experience, tests of face validity, in many dimensions, is extremely valuable in uncovering the most serious errors. It is a prerequisite, however, that the model be well documented and that it be easy for experts to view its behavior (e.g., through interactive post processing graphics rather than fixed hard-copy outputs).

Methods of face-validity testing depend heavily on such things as the following:[32]

- Having a good set of baseline cases (standard scenarios) with which the reviewers are familiar

- Displays of *aggregated* behavior (e.g., total divisions deployed in theater vs time or average divisional loss rates when in combat vs time)

- Highly organized and comprehensible overviews of model approach, assumptions and parameter values (more generally, good documentation is essential; see Annex B for more discussion of documentation)

- The ability to respond quickly to spot-check requests (e.g., "What did you assume for the value of ...?" and "What does the plot of ...vs time look like?" and "Show me, in code, the algorithm (or rules) you used for...")

- The ability to do additional spot-checking cases upon demand (e.g., "Let's see what happens when you assume the B-1's ECM doesn't work.")

The dangers of depending only on face validity are obvious, but they can be mitigated if the effort to do face-validity checks is broad enough, includes opportunities for spot checking in depth, is accomplished with reviewers having a range of backgrounds, and mixes review of "inputs" (model structure, assumptions, etc.) and "behavior." One reason such testing is so valuable is that poorly done models often fail immediately, whereas well done models are the result of serious and professional efforts in which testing and validity-related discussions are an everyday way of life for developers. Given such efforts, intensive review sessions can cover a great deal of ground quickly because the developers are "on top of the problem" and have organized information well.

Detailed validation efforts must depend primarily on module-by-module testing during development and on special meetings to examine critical modules in depth. It is seldom possible with large military models to do anything like comprehensive testing or evaluation of complete multi-module systems.[33]

Special Issues With Knowledge-Based Models.

Knowledge-based models such as rule-based or algorithmic and rule-based decision models representing, e.g., military commanders or operators of air defense systems, raise special issues because in most cases they cannot in principle be validated in the sense of being favorably compared with "the real system." Instead, they must be evaluated on grounds such as whether they faithfully represent the knowledge of relevant experts, whether they are logical, internally consistent and consistent with various physical and logical constraints, and so on.[34] They can in some cases be falsified by real-world experience in which other variables proved to be critical, but ambitions must be limited. Further, there is a wealth of information to the effect that experts often give misleading testimony about what they would do in various circumstances and about the way in which they reason—not because they *intend* to mislead, but because they have only a limited understanding of their own cognition. For example, when being interviewed experts might describe a highly rational process of making decisions, but in the heat of actual operations—with uncertainties, fatigue, and time pressures all being factors—their behavior might reduce to the simplest of patterns, some of them "irrational" from the viewpoint of a decision theorist. To make things worse, most experts have never encountered many of the situations for which we may be asking them to predict behavior. Thus, they are not *really* experts in the same sense that an experienced internist is an expert on childhood diseases.

It follows from this that efforts to validate knowledge-based models, notably behavioral models of various types, including decision models, must depend much more heavily than one might like on combinations of theory, logic, and spotty expressions of expert opinion.[35] It is essential that efforts to build such models be highly organized and that appropriate testing methods be developed. This is an understudied field, but some relevant methods that have been applied in a number of domains are described in Veit, Callero, and Rose (1984) and Veit (forthcoming). These involve developing rigorous factorial designs for comparing model behaviors with behaviors of relevant experts, preferably in circumstances approaching those that would be encountered in the real world, but perhaps in war games as a next-best choice. Another valuable empirical approach is to observe experts performing in field exercises. This can usefully supplement interview data and theoretical analysis by bringing in, to some extent at least, aspects of behavior under stress and the fog of war.

6.3.5. Methods of Accreditation

There are various organizational approaches to accreditation, but this subject is best discussed in the next section.

6.4. A DYNAMIC VIEW OF VV&A

Overview

Figure VI-3 shows a dynamic view of VV&A that emphasizes evaluation and accreditation of a model in the context of a specific study.[36] The importance of context is emphasized by putting the analytic plan in the center. It is here one starts—knowing of course, the purposes of the analysis. Provisional accreditation for a *class* of applications could emerge from a similar chart, but I will not deal with that further in this study.

When evaluating a model for a specific application, chances are that the model is an adaptation of a previous model that has been subjected to some degree of VV&A or that the model has been subjected previously

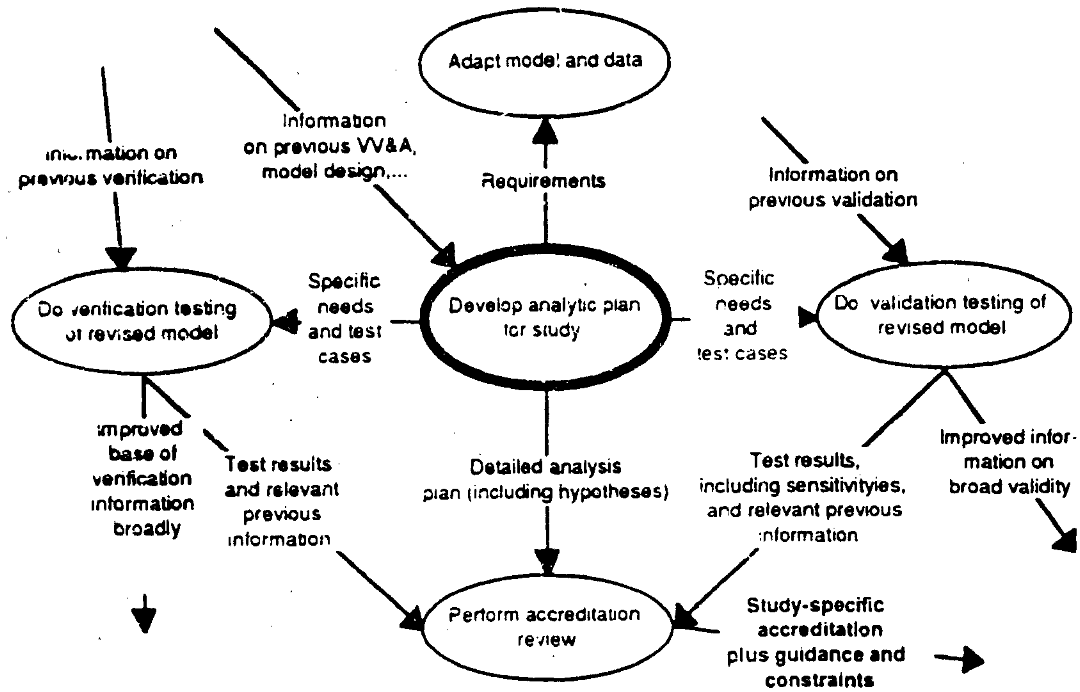

FIGURE VI-3. VV&A as a Continuing Process Sensitive to Context (Process starts at the center)

to considerable "general" VV&A without the benefit of study-specific information.[37] Thus, the new round of VV&A shown in **Figure VI-3** draws on previous information (see arrows coming in from top left). Most importantly, however, it depends heavily on the study-specific requirements and test cases. In practice, relatively complex combat models (or most other models used in policy analysis) are never *fully* tested and unconditionally accredited. Testing can still be extensive and sophisticated for the purposes of evaluating the model and its data in the context of a specific analysis. That testing is the basis for study-specific accreditation, but it also adds to the base of VV&A information that will be used in the next iteration for a new application (see outward arrows on bottom left and center right). One feature of **Figure VI-3** (bottom right) is especially important and unusual. This is its reference to constraints and guidance as outputs of the accreditation process. Since the most stringent review of an analytic organization's work usually occurs within the organization itself, one may think of "accreditation" as being the result of management reviews of the sort that should occur early in a project's life, before the project's work is reported, and, if possible, at least once in between. The result of such a review might take the following form (think of this as the summary conclusions of the relevant manager, who need not be a government official):

> On balance, our conclusions are:
>
> (1) The analytic plan appears to be sound.
>
> (2) The model and data base for carrying out the plan appear to be sound.
>
> (3) Consistent with the improved plan, however, no conclusions should be drawn regarding..., because the analysis cannot support them. Further, in drawing conclusions on..., it is essential that they reflect parametric variations on the following key variables over the ranges discussed in the review. Recipients of the analysis must understand the considerable uncertainty associated with....
>
> (4) Further, recipients of the analysis must be reminded of the following basic assumptions of the approach, which appear reasonable, but which also establish limitations on its significance:...

In this depiction *there is no all-or-nothing blessing of the model—even for a specific study*. Instead, the accreditation is conditional upon the analysis plan itself, which includes the proposed logic to establish conclusions. Further, the accreditation process often results in changes of the analytic plan itself (and changes in the model leading to another round of verification). This iteration is merely implicit in **Figure VI-3**.

In concluding that a model could reasonably be used for the purpose at hand, the accrediting authority might be drawing on highly study-specific information and pondering in some detail precisely what function the model itself is serving (see Hodges and Dewar, 1992 for a list of such functions and related discussion).

One can imagine judgments such as the following being made as part of the accreditation decision and explanation:

> The model is suitable here (e.g., in a war game being used for higher level education and training). Realistically, it is being used primarily as an organizing device, as a kind of bookkeeping mechanism. The results of the analysis depend most sensitively on the human command-control decisions, including operational strategies. The model's treatment of attrition is fairly crude, but as you have shown with your sensitivity analyses, the attrition model is not the limiting factor.
>
> The model is quite suitable here, despite its exceptionally simple treatment of close combat.
> The results depend primarily on the air-to-ground effectiveness of U.S. air forces, given air supremacy, and the time required for us to achieve that supremacy. You

have a rather detailed and credible treatment of both air-to-ground effectiveness as a function of circumstance and of the suppression of air defenses (SEAD).[38]

You must be kidding. The model can't possibly be used to infer conclusions about the proper mix of tank and artillery units, because it bases ground combat attrition on some aggregation expressions that treat MLRS as merely one contributor to an overall firepower. Chances are the model will conclude something like "all we need to do is buy MLRS batteries and disband the rest of the army." That would be fine if battle were just a matter of firepower.

Yes, I know that you think you have a highly sophisticated model of ground combat, but it is not adequate for this study. As it stands, ground forces are unintimidated by air forces, and can maneuver just as quickly with or without enemy air forces attacking them, except to the extent that air forces can destroy whole units. I don't believe this for a moment. Air forces can disrupt and delay, and thereby greatly affect maneuver and tempo generally. Go back to the drawing boards—and read some history on the Battle of the Bulge, especially the part after the weather cleared.

Your model seems fine so far as it goes, covering attrition and movement processes, but it treats operational strategy as input data and doesn't allow adaptation. That leaves out the most important part of force employment. Good forces and bad strategy lead to bad results (see, e.g., Davis and Hillestad, 1992).

An important point to be made here is that the same model might be good for some force-composition or force-structure studies and altogether inappropriate for others. Thus, attempting to accredit a model for whole classes of studies can readily lead to bad decisions. It would therefore seem appropriate to introduce and use the concept of *provisional accreditation* (suggested to me by Clayton Thomas), which would be used in the context of concluding that "This model (and its data base) is a reasonable candidate for use in this kind of study. Go ahead and flesh out the analysis plan and let's then see whether the plan makes sense and the model will indeed be adequate." This emphasizes yet again that it is the analysis, study, or other application that should actually be "accredited."

6.5. ESTABLISHING A VV&A REGIME WITHIN AN ORGANIZATION

6.5.1. Prefacing Comments

In thinking about VV&A and about how to improve its practice in organizations, it is important to recognize that VV&A should not be seen as a separate and segmentable enterprise—i.e., an additional duty or task—but rather as an inherent part of the analytic process from the time of initial design to the time of particular applications. Validation is *central* to the scientific process that good analysis seeks to emulate. I raise these matters here because VV&A is not always viewed in this way. Indeed, there are many considerations that undercut attempts to make analysis "scientific." For example, models are often tools of advocacy; further, data bases are often tightly held for both security reasons and information-is-power reasons. As a result, there are significant *disincentives* for organizations to evaluate their models and data as harshly as they might if they were physical scientists attempting to unravel the secrets of the universe. It is therefore a significant challenge for analytic organizations to rise above these problems and instill and maintain a sense of professionalism and "scientific method." This is a continuing challenge, not one that can be addressed once and for all (see also Hughes, 1989, pp 10 ff). With this background, then, let us examine how an organization might take on the challenge.[39]

6.5.2. Considerations

Establishing a VV&A regime must first be recognized as involving all of the standard challenges associated with organizational change and learning. Simple decrees have very limited and short-term value. Instead, one must think in terms of such matters as:

- Creating and communicating a *vision* of professionalism that treats VV&A as inherent to good work and something to be done continuously rather than merely in occasional painful and unrewarding crash efforts.

- Developing associated policies and procedures, and assuring that there are early examples for everyone to see of how these will be implemented in practice and what will be accomplished.

- Bringing members of the organization into the problem so that they participate in developing aspects of the general policies and many of the procedural details —thereby assuring proper tailoring to the organization's particular culture.

- Establishing the uncomfortable principle of *independent* review, for at least critical features of the work, even though the tendency within organizations is usually to assume that internal review is quite adequate and that the call for independent review is insulting and a potential waste of time.[40]

- In all of this, having both long- and short-term views and plans, with short-term efforts being designed in part to illustrate what is intended on a continuing basis for the long term.

- By distinguishing short- and long-term plans, assuaging fears about unreasonable new demands being added immediately to project burdens.
- Assuring that those contributing to the changes are properly recognized and rewarded.

Many aspects of this challenge can be helped by having concrete examples to use as case histories that everyone reads. An important part of the continuing MORS effort on VV&A is to develop and, if possible, to publish such histories.

6.5.3. Using the Framework

Against this general backdrop of challenges, I suggest using the material of this study as follows:

- Use the definitions and related discussion to communicate the fundamental issues of VV&A.

- Use the taxonomy of VV&A methods (**Figure VI-1**) to broaden perspectives, break down biases, and help establish short-term and long-term plans. In the long-term plan, for example, one might want to use *many* of the validation techniques mentioned, but that would require scheduling and finding support for tasks, or even whole projects, for work that would not ordinarily be done at all (e.g., comparisons with experiences in field maneuvers or large-scale exercises). Thus, the taxonomy should be used primarily as a *checklist*.

- Use the dynamic view of VV&A (**Figure VI-2**) to frame the issues in a realistic, technically solid, and non "political" way. Use it also to develop detailed work schedules for projects—setting aside adequate time for iterative reviews and follow-up model adaptation and testing. Use this view of the problem to highlight the substantive role of accreditation (as distinct from the more political role emphasized by cynics) and its intellectual relationship to traditional guidelines on how to run analysis projects, guidelines that apply also in many ways to applications such as support of exercises and development of decision aids.

- When identifying VV&A requirements for a particular analysis, explicitly consider the costs of fulfilling those requirements. Then, either assure that the requirements can be met by making available the necessary resources and calendar time or adjust the analyst plan (or claims made about the analysis when concluded).[41]

- Take seriously the discussion of how special measures need to be adopted in evaluating knowledge-based models and other models for which hard data is lacking. Use the examples provided here and develop important distinctions for the problems at hand.

- Use **Figures VI-1, VI-2** and related discussion to explain to sponsors how VV&A plans are consistent with a comprehensive view of the subject, drawing also upon other published materials such as Sargent (1987) and methods used by Martin Marrietta (1990). As part of this, focus sponsors and accrediting authorities (usually the same individuals) on the view of accreditation that encourages them to provide intellectual guidance, not merely a "yes" or "no" decision. And, as part of this, emphasize the need for VV&A activities to be adequately supported and scheduled realistically over time.

Finally, let me mention again that the examples in this study emphasize applications in which models are used for analysis. Many readers will wish to develop analogous examples for their own applications, which may to training, education, operational decision aids or other matters. While the basic framework should hold up, the detailed criteria for judging models is very application dependent.[42]

ANNEX A - ON SEPARATING CONCEPTUAL MODELING AND PROGRAMMING

In a classical ideal with which I long had sympathy, the design and review of models (sometimes called conceptual models) precedes programming.[43] One develops the conceptual picture and lays out the theory and algorithms formally, thereby creating machine- and language-independent specifications (see, e.g., **Figure VI-A-1** from Sargent's work, which remains useful even if my arguments here are accepted). Implementation as a program then proceeds, but its details depend on hardware, software, local practices, and other factors.[44] In this ideal, substantive discussion should focus on the model, not the program. This ideal has much to recommend it, because enormous confusion is caused by having problem formulation shaped and described in terms peculiar to particular languages or computer systems.

In practice, however, the ideal breaks down for both good and bad reasons. The principal bad reason is that many organizations lack the discipline to enforce serious design before allowing programmers to write code; the results are predictable: incomprehensible models that are merely implicit in long and complex computer code.

The good reasons have to do with technology and the changing ways in which worker, even workers with a theoretical ben, go about their efforts. It is becoming increasingly possible and attractive to work largely at the computer rather than with pencil and paper—even for constructing top-down conceptual designs. Second, some of the computer tools for doing so blur the distinction between design and programming, because when one creates the initial design elements (e.g., variable names, data structures such as objects, functions, and diagrams), the results automatically generate corresponding program elements (see Annex B). Third, with some high-level languages, it is as easy for reviewers to understand and comment upon algorithms expressed as computer code (or related diagrams) as it is for them to do so in a paper-and-pencil mode.[45] Fourth, advanced tools such as *Mathematica*™ now make it possible to solve equations symbolically on line, which enhances the design process. And, lastly, statements of the conceptual model often underspecify the problem, resulting in programmers filling in and thereby having much more of a role in defining the "real" model than was intended. In some respects, it is only realistic to force model designers to address explicitly what they might otherwise tend to assume are mere implementation issues (e.g., time steps, control flow in procedural problem-solving approaches, and whether to organize around data structures or processes).

A related issue here is that of prototyping. In the last decade workers have come to appreciate the efficiency of rapid prototyping as a mechanism for helping designers understand the problem for which they are tasked to build models. In practice, it is common for even first-rate modelers and analysts to misunderstand major elements of the problem until they have actually built something and worked with it. While preliminary design is neces

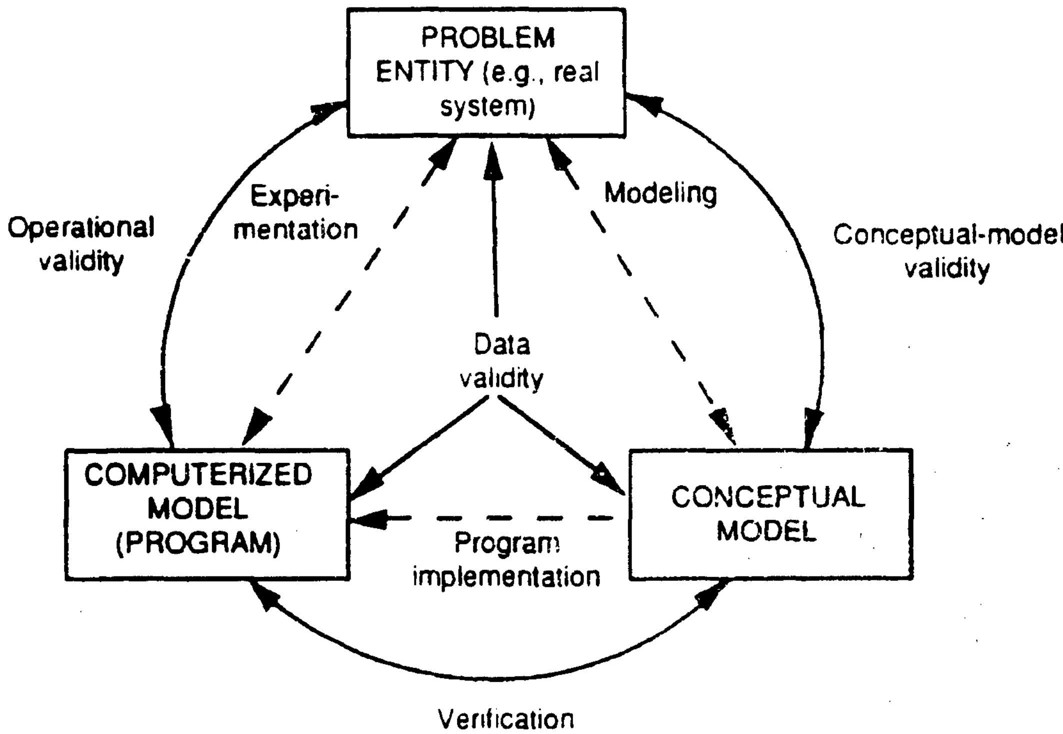

FIGURE VI-A-1. An Idealized Separation of System, Model and Program

sary, it is seldom sufficient and those with modern software tools tend strongly to recommend highly iterative development that exploits prototyping and the discovery process as an inherent part of high quality work, not something to be apologized for.

While I continue to recommend separating model design from design of detailed implementation, and while I still believe it is desirable for many aspects of a model to be reviewed away from the computer context, which tends still to encourage a linear line-by-line view and inelegant solution techniques, the original ideal is now, in my view, obsolete. It is a major challenge for developers to create new operating procedures that will maximize benefits of computer environments while maintaining an appropriate separation of model and implementation detail.

ANNEX B - DOCUMENTATION, HIGH-LEVEL COMPUTER LANGUAGES, AND MODERN MODELING AND ANALYSIS ENVIRONMENTS

6.B.1. DOCUMENTATION

A prerequisite for VV&A is documentation, but many DoD combat models are inadequately documented. To improve this situation, it is important to know what constitutes good documentation. The DMSO's Applications and Methodology Working Group discussed this at some length in 1991, drawing heavily on experience of the participants, many of whom had actually developed large models and/or evaluated them in detail. It agreed that the following guidelines are especially important:

- Distinguish model from program (i.e., describe the conceptual model in terms that are language independent and focused on the underlying concepts and relationships)
- When appropriate, describe model in object-oriented terms, even if the implementing program is not object oriented[46]
- Require high-level designs describing motivation, rationale and basic assumptions, plus:
    - Hierarchical top-down structures (where hierarchies apply) and data-flow diagrams to show how inputs get transformed into outputs
    - Meanings of variables (input to data dictionaries)
    - Logical or algorithmic detail on selected key modules
    - Structured and commented source code, even though this cannot replace documentation, especially higher level documentation
    - Program and interface documentation and illustrative-scenario "walkthroughs"

Distinguishing the model from the program is important in sharpening and communicating concepts, even if the arguments of Annex A are accepted. Programmers often talk about pointers, memory, stacks, arrays, and other constructs having nothing to do with military phenomenology. Documentation and reviews of model content should instead focus on phenomenology.

One important element of good documentation is often overlooked: including the procedures and results of any previous VV&A efforts conducted during development or applications. This can be exceptionally useful.[47]

There are limits to how much documentation can be squeezed out of money-limited projects. The most important documentation consists of "High Level Designs," which are top-down in character with an emphasis on structure. These should also define key variables, provide appropriate diagrams showing, e.g., information flow and control flow, and provide logical or algorithmic detail on key submodels. It is less important, and may even be inappropriate, to document details of much of what constitutes a complex combat model, since those details are often book keeping methods best understood at the level of the code itself. The code, however, should be well structured and commented. Another major element of documentation is information on how to use the program and its interfaces. This is often best done by providing a step-by-step discussion of how one runs and analyzes a test case (i.e., a walk through of a representative application in a given scenario). Commercial software tools often have excellent "walkthrough" documentation.

Taken together, then, there is need for documentation on the model, the program, and its use. Increasingly, on-line documentation is becoming especially important for procedural information.

Finally, note that documentation methods should be changing, and that should be reflected in work on comprehensive environments.

6.B.2 HIGH-LEVEL LANGUAGES AND ENVIRONMENTS

The phrase "high-level language" is ambiguous, because there are multiple dimensions along which to measure. SIMSCRIPT™ was one of the first high-level languages designed for simulation. It was high level in such respects as providing tools making it easy to construct simulations. It also had mechanisms to force good programming practices such as writing an overview of the model, using descriptive identifiers, and exploiting class concepts. In more recent times, spreadsheet languages such as EXCEL™ may be considered very high level in the sense of having user friendly interfaces and a myriad of predefined functions. At the same time, spreadsheet programs are usually the antithesis of structured programming, because the approach taken by the novice is to organize by spreadsheet cells and use the equivalent of many GO TO statements producing "spaghetti code." Further, complex spreadsheet programs based on the systematic use of macros are no more intelligible than those of other languages such as BASIC, and arguably less so.

Against this background, RAND has been developing high-level languages that emphasize using relatively natural language for key words and that exploit the cognitive effectiveness of table structures for organizing both information and logic. RAND now has seven years' experience with RAND-ABEL™, which has been used to write hundreds of thousands of lines of code. The applications have ranged from decision models (e.g., those of a simulated theater commander) to combat models (e.g., attrition and movement processes for combat taking place on a network). It has consistently proven possible to have group reviews of major portions of these models by working directly with code, even though many of the participants have not been serious programmers. Errors have been discovered at a glance, and complex logic has been discussed as a group. Most of this has been possible because of the table structures, which should be developed in other languages as well.

In current work, RAND is developing an object-oriented version of RAND-ABEL, called Anabel.[44] This will extend the effort to exploit two-dimensional structures of many kinds (e.g., decision tables,

tables of orders, and adjudication tables) and will also include numerous self-documenting features, including the use of hyper media. Our belief is that model documentation will not improve greatly by virtue merely of managers cracking whips. Instead, there is both need and opportunity for technology to help. Similar ideas are being pursued at many levels by a variety of researchers, including some who are contemplating the use of expert systems to help choose and use verification and validation tools (see, e.g., Ören, 1986 and Sargent, 1986). In addition, researchers are developing a variety of excellent graphical tools, some of them capable of generating code directly. The Systems Dynamics programs Stella™ and iThink™ are especially notable here. Plans call for a variety of such tools to be used with RAND's Anabel, building on tools recently developed by Larry McDonough and Richard Hillestad. One, called Mapview, allows workers readily to create objects and emplace them on maps. The results of what they do with the graphical interface generate code. Similarly, a tool called the Activity Sequence Editor (ASE) allows workers to develop state-transition diagrams for object-oriented programs, and to have the results of those diagrams generate code. All of this facilitates documentation and VV&A, because many aspects of model design are best seen graphically, and because the tight linkage between diagrams and code avoids the traditional problem of documentation lagging the reality embedded in the code itself. Despite the progress, however, there is a great deal to be done in this general subject area.

6.B.3 A THREAT TO ADVANCEMENTS

Progress in developing and disseminating advanced modeling and analysis methods and tools, including many that would facilitate VV&A, will be adversely affected if the DoD attempts to force all modeling activities into a single structure and language, such as Ada in particular. Such a policy would hinder efforts to exploit the rich selection of commercial products that exist and are emerging. It would also hinder efforts to develop advanced tools, many of which are most readily developed within existing computer environments (e.g., Unix and MacIntosh). The motivation for commonality is understandable, and the desire for greater reusability and interoperability of software is laudable, but the requirement for a single language is misplaced. *High degrees of reusability and interoperability can be accomplished with standards that are language independent.* Indeed, that is what makes "open architectures" feasible and important. Ada is a powerful language that can greatly contribute to the management and control of software development in many projects, but it is much less suitable for prototyping, or for models that will continue to change and that deal with highly uncertain phenomena. For such models there is a high premium on, e.g., interactiveness, flexibility, clarity, explanation capabilities, and easy connectivity to commercial tools.

BIBLIOGRAPHY

Anderson, Robert H., Steven C. Bankes, Paul K. Davis, H. Edward Hall, and Norman Z. Shapiro (forthcoming), *Toward a Comprehensive RAND Environment for Computer Modeling, Simulation, and Analysis.*

Bonder, Seth (1984), "Summary of a Verification study of Vector-2 with the Arab-Israeli War," in Reiner Huber (editor), *Systems Analysis and Modeling in Defense: Developments, Trends, and Issues*, Plenum Press, New York, 1984.

Brewer, Gary and Martin Shubik (1979), *The War Game: A Critique of Military Problem Solving,* Harvard University Press.

Davis, Paul K. (1990), *An Analyst's Primer for the RAND-ABEL™ Programming Language,* RAND, N-3042-NA.

Davis, Paul K. and Donald Blumenthal (1991), *The Base of Sand Problem: a White Paper on the State of Military Combat Modeling*, RAND, N-3148-OSD.

Davis, Paul K. and Reiner Huber (1992), *Variable-Resolution Modeling: Motivations, Issues, and Principles*, RAND, N-3400-DARPA.

Davis, Paul K. and Richard Hillestad (1992), "Using Simulation in the Education of General Officers," *Proceedings of the Summer Simulation Conference.*

Davis, Paul K., Steven C. Bankes, and James P. Kahan (1986), *A New Methodology for Modeling National Command Level Decisionmaking in War Games and Simulations*, RAND, R-3290-NA.

Defense Modeling and Simulation Office (1991), *Final Report of the Applications and Methodology Working Group,* December, 1991.

Department of Defense/Defense Modeling and Simulation Office (1992), *Defense Modeling and Simulation Initiative,* May 1.

Dupuy, Trevor (1987), *Understanding War,* Paragon House, New York.

Elzas, Maurice S. , Tuncer Ören, and Bernard Zeigler (editors), *Modelling and Simulation Methodology in the Artificial Intelligence Era*, North-Holland, Amsterdam, 1986.

Gass, Saul (1983), "Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis," *Operations Research, Vol. 21, No. 4, July-August, 1983.*

Helmbold, Robert L. (199;), *Point Paper: CAA History Activities, 1980-1990*, U.S. Army Concepts Analysis Agency, Bethesda, MD.

Hodges, James and James Dewar (1992), *Is It You or Your Model Talking? A Framework for Model Validation*, RAND, R-4114-RC/AF.

Hughes, Wayne, editor (1989), *Military Modeling, Second Edition*, Military Operations Research Society, Alexandria, Virginia.

Klir, G.J. (1989), "Inductive Systems Modeling: An Overview," in Elzas, Ören, and Zeigler (1989).

Martin Marietta (1990), *Confidence Methodology Guide*, Third Edition, Final, National Test Bed Technical Report NTB-237-022-06-02, August 15. Prepared for Strategic Defense Initiative Organization, Washington, D.C.

McQuie, Robert (1987), *Historical Characteristics of Combat for Wargames (Benchmarks)*, Army Concepts Analysis Agency, CAA-RP-87-2.

Military Operations Research Society (MORS) (1989), *Human Behavior and Performance as Essential Ingredients in Realistic Modeling of Combat—MORIMOC II*, Alexandria, Virginia. (Proceedings of a conference held 22-24 February, 1989).

Miser, Hugh J. and Edward S. Quade (1988), *Handbook of Systems Analysis: Craft Issues and Procedural Choices*, Elsevier Science Publishing Co., New York, New York.

Ören, Tuncer (1986), "Artificial Intelligence in Quality Assurance of Simulation Studies," in Maurice S. Elzas, Tuncer Ören, and Bernard Zeigler (editors), *Modelling and Simulation Methodology in the Artificial Intelligence Era*, North-Holland, Amsterdam, 1986.

Ören. Tuncer (1989), "Bases for Advanced Simulation: Paradigms for the Future," in Maurice Elzas, Tuner Ören, Bernard Zeigler, editors, *Modelling and Simulation Methodology*, North-Holland, Amsterdam, 1989.

Pace, Dale K. and Dennis P. Shea (1992), "Validation of Analysis Which Employs Multiple Computer Simulations," *Proceedings of the Summer Simulation Conference*, pp. 144-149, Society for Computer Simulation.

Palmore, Julian (1992), "Analysis and Verification and Validation of Complex Models," *Proceedings of the Summer Simulation Conference*, pp 139-144, Society for Computer Simulation.

Sargent, Robert (1986), "An Exploration of Possibilities for Expert Aids in Model Validation," in Elzas, Ören, and Zeigler (1986).

Sargent, Robert (1987), "An Overview of Verification and Validation of Simulation Models," *Proceedings of the 1987 Winter Simulation Conference*, Society for Computer Simulation.

Sargent, Robert E. (1987), "An Overview of Verification and Validation of Simulation Models," *Proceedings of the 1987 Winter Simulation Conference*, Society for Computer Simulation.

Shapiro, Norman, H. Edward Hall, Robert Anderson, Mark LaCasse, James Gillogly, and Robert Weissler (1988), *The RAND-ABEL Programming Language: Reference Manual*, RAND, N-2367-1-NA.

Shlapak, David and Paul K. Davis (1992), *Possible Postwar Force Requirements for the Persian Gulf: How Little Is Enough?*, RAND, N-3314-CENTCOM/JS.

Stockfisch, John A.(1975), *Models, Data, and War: A Critique of the study of Conventional Forces*, RAND, R-1526-PR.

Thomas, Clayton (1983), *"Verification Revisited,"* in Hughes (1989). In the original 1983 edition of Hughes' book, Thomas used "verification" in the sense that is now applied to "validation." He changed his usage for the 1989 edition cited here.

U.S. Army (1992), Army Regulation 5-11, Army Model and Simulation Management Program, June.

U.S. General Accounting Office (1980), *Models, Data, and War: A Critique of the Foundation for Defense Analyses*, GAO/PAD-80-21. Washington, D.C.

U.S. General Accounting Office (1987), *DoD Simulations: Improved Assessment Procedures Would Increase the Credibility of Results*, GAO/PEMD-88-3, Washington, D.C.

Veit, Clairice (forthcoming), *Methods for Developing Validated Subjective Measures to Incorporate in Simulations*, RAND, R-4209-A/DDR&E.

Veit, Clairice, Monti Callero, and Barbara Rose (1984), *Introduction to the Subjective Transfer Function Approach to Analyzing Systems, RAND*, R-3021-AF.

Waterman, Donald A. (1986), *A Guide to Expert Systems*, Addison-Wesley, Reading, Mass.

Williams, Marion L. and James Sikora (1991), "SIMVAL Minisymposium—a Report," *Phalanx*, Vol. 24, No. 2, June.

Zeigler, Bernard (1976), *Theory of Modelling and Simulation*, John Wiley and Sons, New York; reissued by Krieger Pub. Co., Malabar, Florida, 1985).

Zeigler, Bernard , *Multifacetted Modelling and Discrete Event Simulation*, Academic Press, London and Orlando, FL., 1984.

Zühtü, Aytaç and Tuncer Ören (1986), "MAGEST: A Model-Based Advisor and Certifier for GEST Programs," in Elzas, et. al. (1986).

ENDNOTES

1. Adapted from a RAND report of the same name, R-4249-ACQ, Santa Monica, CA 1992.

2. Standard references of this sort include Brewer and Shubik (1972), U.S. GAO (1980), and U.S. GAO (1987), which contains an extensive bibliography. One of the most famous essays on the subject is Stockfisch (1973). Davis and Blumenthal (1991) examines broader issues and argues that many problems in combat modeling stem from failure of the military community to think in terms of nurturing a robust military *science*. The near-exclusive emphasis on models as mere tools has been an obstacle to seeing some models as *theories* that need to be developed, tested, and evolved scientifically.

3. Another useful reference is Williams and Sikora (1991), which provides a snapshot view of continuing work on VV&A by the Military Operations Research Society (MORS). Readers may wish to check for updates in the newsletter *Phalanx*. MORS hopes to publish a book on VV&A sometime in 1993. This study may contribute to that effort.

4. Some sources define "simulation" differently—as the operation or exercise of a model, or as a method of implementation.

5. As an example, consider a model predicting the damage expectancy for a set of hard targets as a function of a bomber's availability, reliability, pre-launch survivability, penetration probability, bomb load, and hard-target kill capability. The bare model provides an intellectual framework, but has little or no predictive value: its predictions are "data driven." Similarly, in idealized knowledge-based systems such as an expert system describing likely decisions of a commander, the bare model may be a general "inference engine" for processing rules, while the content of the model resides entirely in the "knowledge base" of rules such as "If we can achieve surprise and if the force ratio is no worse than...Then we shall...".

6. Another way in which the classical distinction between model and data has broken down is with the introduction of highly interactive computer languages, which make it possible for users to change many equations and structures in the computer code as easily as they can change the data value used for the gravitational constant. The most familiar example of this is in spreadsheet programs, but other examples include BASIC and RAND-ABEL.*

7. A related issue here is establishing that the numerical procedures used are not introducing cahoas effects. Palmore (1992).

8. Articles on software engineering sometimes use terms such as "rigorous audit" or otherwise convey the impression of verification requiring complete testing over all computational "paths." Except at the level of relatively small modules, however, such review and testing is usually not feasible. Thus, there is a premium on designing a doable set of tests that will be likely to uncover the most serious problems.

9. Some of the following discussion draws on review comments by Mr. Dennis Shea of the Center for Naval Analyses. See also Pace and Shea (1992).

10. As discussed later in the study, there is a number of modern techniques that can automate or otherwise assist a good deal of verification testing. Many depend on the existence of a data dictionary that is part of the language or environment, not a mere repository of comments.

11. As of April 1, 1992, the MORS group concerned with VV&A was using as a working definition: "The process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model."

12. Other workers sometimes refer to structural validity vs output validity. In that breakdown, output validity includes both descriptive and predictive validity.

13. The subject of resolution is complex and analysts often need to work with models with different resolutions which, ideally, are consistent in the aggregate. See Davis and Huber (1992) for related discussion.

14. One effort to assess descriptive validity is described in Bonder (1984), which examines the ability of the Vector-2 model to reproduce the battle for the Golan Heights in the Yom Kippur War.

15. Other decompositions are possible. Based on discussions at the MORS SIMVAL II meeting, Dr. Dale Henderson of Los Alamos National Laboratory decomposes the space of validation activities into five dimensions: (a) the techniques used (e.g., Delphi vs quantitative comparisons), (b) the basis of truth used (e.g., historical data vs. results of more detailed simulations), (c) the applications intended for the model, (d) the degree of composition at which testing occurs (e.g., on primitive modules vs higher-level subsystems or a complete integrated system), and (e) the depth of the validation effort (e.g., surface-level or face-validity testing). The principal point is that validation activities are multidimensional rather than rank-ordered or hierarchical.

16. Prehistoric man presumably "knew" that the sun would come up every morning and that there was a cycle of progressively longer and then progressively shorter days. He presumably counted on this model long before there was any understanding of astronomy.

17. Most of our "stochastic processes" are at their root deterministic; the problem is our uncertainty about initial values and interactions with other processes, which causes us to treat them as stochastic.

18. In practice, application-specific accreditation usually depends (and *should* depend) on an assessment of the people and organization using the model, not merely the model itself. Indeed, one can argue that it is more important to "accredit" (or at least to assess) people and organizations than the tools they use.

19. A classic example of this is use of silo hardness, measured in psi. Many strategic-nuclear analyses have been conducted using silo hardness, even though the phenomenology of silo destruction is complex and requires something more sophisticated, such as a vulnerability number approach that accounts for effects of both static and dynamic pressures. Analysts can nonetheless get by with computer programs or analytic models using hardness, because they do offline calculations to derive the effective hardness of silos to the weapon yields of interest.

20. One can argue that the issue of clarity applies more to the *study* or other application than to the model itself, but those interested in the clarity (and reproducibility) of studies are usually driven toward seeking clarity of models as well. While it is true *in principle* that analysis with black-box models can be clear, given enough sensitivity testing, my own experience is that depending on such an approach is usually a recipe for disaster.

21. As an example here, if one knows the detailed application, one can develop tests of the integrated system using relevant parameter values. Without such knowledge, full-system testing may be extremely difficult because of the number of possible combinations possible.

22. See Zühtü and Ören (1986), Sargent (1986), and Ören (1986) for discussion of ambitious ideas going beyond the examples given here.

23. Most of these techniques require an "active data dictionary, " which is a data base of information on the model's data—e.g., information on type, format, acceptable values and meaning. Except for "meaning," the information can be used automatically to check source code and data values.

24. Hodges and Dewar (1992) argue that failure to appreciate this reality has been a fundamental source of difficulty in the continuing discussions about validating military models. They argue that the word "validation" should be reserved for predictive models that can be rigorously tested , and that other types of model evaluation should be developed as a function of how the models are to be used (e.g., as bookkeeping devices in a human war game, as decision aids, and as devices to stimulate hypotheses).

25. It is not uncommon for "theories" to be expressed in ways that make it impossible to disprove them. Good science, by contrast, insists that theories be falsifiable. Indeed, scientists go to considerable lengths to define experiments that stress their theories as much as possible.

26. As an example of where military science might enter, consider the many theater-level models over the years in which air forces for close air support and battlefield interdiction have not been concentrated in time and space, thereby diluting their potential effect on the other side's ground-force maneuver and ignoring the importance of concentration and coordination to military art generally and to survival and effectiveness of those air forces specifically. As another example, consider the common failure to represent adequately the suppressive effects of artillery. There are models, of course, which handle both of these issues relatively well, but many military models have grossly misrepresented the phenomena, often without justifying their simplifications through auxiliary calculations. Detecting such problems is arguably a matter of "science," not logic or analytic rigor.

27. It is striking to note that theoretical evaluation is commonly (almost always) omitted from discussion of validation methods. It is most assuredly not the same as "logical verification" or "logical testing." My own sense is that the omission is another symptom of military modeling suffering from not being part of a military science. It has perhaps been overly influenced by mathematicians and programmers, without the emphasis on phenomenology that scientists are supposed to bring to the table (but scientists can also be beguiled by simplistic but elegant mathematics). An important role for military officers, including retired general officers serving as consultants, is to insist that modelers pay more attention to the *real* phenomena. They must demand more military science if the models are to be faithful to their needs.

28. In MORS work the distinction has been drawn between "output validation" and "structural validation." One can map the activities of Fig. 3.2 into these terms, but not neatly. Theoretical evaluation includes both structural validation and testing behavior (outputs) in various special cases that are understood with prior theories or for which there exist solid empirical data. Empirical evaluation in Fig. 3.2 relates to output validation in MORS terms. "Other comparisons" in Fig. 3.2 involve both structural and output validation. For example, comparisons to expert opinion and doctrine can look both at assumptions and output.

29. This view treats validation as a matter of degree. Hodges and Dewar (1992) take a different approach.

30. As one reviewer of this report noted, "doing something" sometimes should mean doing the best analysis possible even though that means *not* using a computer model that sponsors and users of the computer model are expecting will be used. This may be logically obvious, but it can be a problem in practice because there are instances in which reference to a well known computer model is thought somehow to confer a sense of validity, legitimacy, or acceptability.

31. MacQuie (1987) is an interesting compilation of historical data to be used in tests of face validity. The Army's Concepts Analysis Agency has a continuing effort to exploit historical data (see Helmbold, 1990 for references).

32. Even more fundamental is the need for professional model development practices emphasizing module-by-module testing by developers as a routine part of everyday work. If more sloppy methods have been followed, face-validity efforts are likely either to fail or be quite misleading.

33. Importantly, much more extensive testing *would* be possible if it were budgeted. It is unusual, however, for military simulation projects to set aside, e.g., 20% of the overall project funds for independent and comprehensive VV&A. In some instances, such testing would be well worth the investment. In many other cases, however, some imperfections are quite tolerable.

34. Some concrete examples here come from a recent evaluation by the Center for Naval Analyses (CNA) of a command and control model. The review asked: (a) Have all the decision nodes been identified?; (b) For each node, has a variable been defined for each factor that could affect decisions at that node?; and (c) For every possible state of each variable at each node, has a rule been developed (e.g., an If/Then statement) and does the rule reflect the judgement of experts?

35. My own experience with knowledge-based models has emphasized theory and logic, with experts being used mostly for spot checking. See, e.g., Davis, Bankes, and Kahan (1986). The textbook concept of using "knowledge engineers" to extract knowledge from experts often does not apply or is less efficient and organized than having a subject-area analyst build a model and then iterate it by talking with experts. For a discussion of the knowledge-engineering approach, see Waterman (1986).

36. This discussion envisions a model being used for an analysis study. However, analogous diagrams could readily be constructed for such other applications as training, education, and operational decision aids. Some readers may wish to do so.

37. There is an issue of balance and complementarity here. Some discussions of VV&A convey the impression that models can be adequately evaluated once and for all, when in reality model appropriateness must be judged in the context of an application. However, studies often occur with time pressures and modest resources, which means that they cannot take on the full burden of evaluating models from scratch and depend on there having been a considerable degree of prior VV&A. While Fig. 4.1 deliberately focuses on VV&A for an application, both that and the broader VV&A are increasingly considered essential (e.g., US Army, 1992). Personally, I would argue that generic V&V is essential, and generic accreditation is potentially useful (and potentially troublesome), depending on organizational sophistication, integrity, and efficiency.

38. In a similar spirit, a colleague and I conducted a study of possible post-crisis defense requirements a few months before the allied offensive against Saddam Hussein, in which we used an extremely simple spreadsheet model using Lanchester equations and aggregated force strengths for ground combat. The reason for doing so was that we observed results of more sophisticated and complex war gaming analysis were driven by a few factors (e.g., air-to-ground effectiveness) that were being obscured by the original level of detail (see Shlapak and Davis, 1992). For other purposes, however (e.g., evaluating *offensive* capabilities), the simple model would have been ludicrously inappropriate.

39. Although not discussed in this study, a major issue is how the DoD can create positive incentives for VV&A. Currently, most of the "incentives" under discussion are in the nature of requirements and threats. The most obvious incentive, however, is money: by budgeting appropriately for serious VV&A, The DoD would quickly find itself receiving first-rate proposals for high-quality testing. The second principal incentive I see is the fostering of an invigorated military science as discussed in Davis and Blumenthal (1991).

40. There is a strongly held view in the larger software community that good VV&A is necessarily *independent* VV&A. Indeed, it is not uncommon to have separate organizations charged with development and VV&A. The motivation here is recognizing that developers often have profound conflicts of interest that undercut VV&A. The pressures include deadlines, cost, the desire to include new and more sophisticated submodels, and the antipathy of workers for the drudgery of extensive testing. An independent tester paid specifically to certify software has, by contrast, other incentives. At the same time, there is substantial evidence demonstrating that "independent testing" cannot usually be conducted in isolation: it is essential for the testers to interact with both developers and

users. Developing appropriate working relationships that balance independence of judgment with cooperation and exchange of information is therefore important.

41. The issue of budgeting for VV&A is fundamental, and the failure to appreciate this probably underlies many of the VV&A problems in the military modeling community.

42. As one example, consider that program planners often think in terms of aggregations that are of little or no value to officers participating in operational exercises. As a result, they need different models. Ideally, the models will be consistent, but that is not always easy (Davis and Huber, 1992).

43. See, e.g., Zeigler (1984), Sargent (1986, 1987), Gass (1982). and Martin Marrietta (1990).

44. As discussed by Julian Palmore of the University of Illinois in an address to the 60th MORS conference in Monterey, California in June, 1992, even details of computer arithmetic can be very important in simulation. Failure to pay attention to such details can produce substantial "structural variance" as manifested, e.g., by peculiar sensitivity results and major changes in results if one shifts from one computer to another. See also Palmore (1992).

45. Separate documentation is still needed for gaining a top-down overview of the model and program. Further, it is virtually essential when the program itself is large. However, the documentation may be out of date or may contain errors that do not exist in the code (and, of course, the code may contain errors not in the documentation). My own view is that future reviews of models should ideally combine reading of documentation for top-down structure and having that documentation, which may also be on line, "point to" critical portions of code that can be examined directly. That will be increasingly feasible with high-level computer languages and environments (see Annex B).

46. One can design a model in terms of objects, attributes, processes, and the like whether or not the programming language has the paraphernalia of objects, messages, methods, and so on.

47. In naval modeling a special need is discussion of how environment is handled in the model.

48. Anabel, the result of ideas by Edward Hall and Norman Shapiro, is being developed as part of a grander scheme for a modeling and analysis environment (see Anderson, Bankes, Davis, Hall, and Shapiro, forthcoming). RAND-ABEL is documented in Davis (1990) and Shapiro, Hall, Anderson, LaCasse, Gillogly, and Weissler (1988).

Appendix A

SELECTED BIBLIOGRAPHY

Ackoff, R. L. 1962. *Scientific Method: Optimizing Applied Research Decisions.* New York and London: John Wiley and Sons, Inc.

Adamson, D. S. Andreani, E. C., Archer, G. W., et al. 1977. *Mathematical Model User's Manual Combined Arms Tactical Training Simulator (CATTS) Device 16A3. Volumes I-IV.* Redondo Beach, California: TRW, Inc. (AD A038 796; AD A038 797; AD A038 798; AD A038 799).

Amdor, Stephen L. 1976. TAC-WARRIOR--A Campaign Level Air Battle Model. In *Proceedings of the 37th Military Operations Symposium (June 1976).* Alexandria, Virginia: Military Operations Research Society (AD C009 755L).

Anderson, Howard C. 1978. Command and Control Considerations Relevant to the Combined Arm Simulation Model (CASM) (U). In *Proceedings of the 39th Military Operations Research Symposium (June 1977).* Alexandria, Virginia: Military Operations Research Society.

Anderson, Lowell B., et al. 1975. *Revised OPTSA Model. Volume I: Methodology.* Arlington, Virginia: Institute for Defense Analysis (LD 35385MA).

Anderson, Lowell B., et al. 1976. *In Maxmin and Minmax Strategies in Multistage Games and ATACM.* Arlington, Virginia: Institute for Defense Analysis (AD A030 246).

Anderson, Theodore Wilbur. 1958. *An Introduction to Multivariate Statistical Analysis.* New York: John Wiley and Sons, Inc.

Bailey, T. J., and Martin, G. A. 1977. *CACDA Jiffy War Game. Technical Manual. Part I: Methodology.* Fort Leavenworth, Kansas: Army Combat Development Activity (LD 39809MA).

Barbieri, W. A. 1973. *Countermilitary Potential: A Measure of Strategic Offensive Force Capability.* Santa Monica, California: Rand Corporation (R-4131-PR).

Battilega, J. A., Blackwell, L. M., and Phelps, M. W. 1977. *An Introduction to Quantitative Strategic Force Analysis.* Denver, Colorado: Science Applications, Inc. (SAI-77-011-DEN).

Battilega, J. A., Grange, J. K., eds. 1978. *The Military Applications of Modeling.* Wright-Patterson AFB, Ohio: Air Force Institute of Technology Press.

Battilega, J. A. 1974. *Selected Mathematical Programming Techniques for Force Allocations Problems*. Denver, Colorado: Science Applications, Inc. (SAI-74-017-DEN).

Battilega, J. A. 1975. *Multi-Goal Optimization With the Arsenal Exchange Model.* Denver, Colorado: Science Applications, Inc. (SAI-75-30-DEN).

Battilega, J. A., and Cotsworth, W. L. 1977. *Multiple Criteria, Strategic Force Analysis, and the Arsenal Exchange Model.* Denver, Colorado: Science Applications, Inc. (SAI-77-058-DEN).

Beach, Earl F. 1957. *Economic Models: An Exposition.* New York: John Wiley and Sons, Inc.

Benson, K. T., and Miller, C. E. 1974. *User's Manual for QUANTO--A Weapon Allocation Code.* Kirtland AFB, New Mexico: Air Force Weapons Laboratory (AD 779 754).

Benson, K. T., et al. 1974. *QUANTO--A Code to Optimize Weapon Allocations.* Kirtland AFB, New Mexico: Air Force Weapons Laboratory (AD 773 801).

Bergstrom, A. R. 1966. Nonrecursive Models as Discrete Approximations to Systems of Stochastic Difference Equations, *Econometrica,* Volume 34, No. 1 (January), pp. 173-182.

Bishop, Albert B., and Stollmack, Stephen. 1968. *The Tank Weapon System. Final Report.* Columbus, Ohio: Ohio State University (AD 850 367).

Blackett, P. M. S. 1962. *Studies of War.* New York: Hill and Wang.

Blackwell, David, and Girshick, M. A. 1954. *Theory of Games and Statistical Decisions.* New York: John Wiley and Sons, Inc.

Blackwell, L. M., et al. 1976. *UNICORN (Version III) Methodology.* Denver, Colorado: Science Applications, Inc. (SAI-76-048-DEN).

Bode, J. R. 1974. *Indices of Effectiveness in General Purpose Force Analysis.* Vienna, Virginia: Braddock, Dunn, and McDonald (BDM/W-74-070-TR).

Bonder, Seth. 1976. *A Summary Descripton of the VECTOR-II Theater-Level Campaign Model.* Ann Arbor, Michigan: Vector Research.

Bonder, Seth, and Farrell, Robert. 1970a. *Development of Analytical Models of Battalion Task Force Activities.* Ann Arbor, Michigan: Systems Research Laboratory, Michigan University (AD 714 677).

Bonder, Seth, and Farrell, Robert. 1970b. *Development of Models for Defense Systems Planning*. Ann Arbor, Michigan: Systems Research Laboratory, Michigan University (AD 715 646).

Bozovich, J. F., et al. 1973. *Mathematical Formulation of the Arsenal Exchange Model (Revision 7)*. Denver, Colorado: Martin Marietta Corporation.

Bracken, J. 1973. *Two Optimal Sortie Allocation Models*. Arlington, Virginia: Institute for Defense Analysis (IDA Paper P-993).

Bracken, J., et al. 1975a. *Quantity-Quality Tradeoffs*. Arlington, Virginia: Institute for Defense Analysis (HQ 74-16712, Study 5-443).

Bracken, J., et al. 1975b. *A Theater-Level Analysis of NATO Aircraft Shelters and Air-Defense Attrition Rates (U)*.
Arlington, Virginia: Institute for Defense Analysis (IDA Study 5-447).

Brode, H. L. 1970. *Height of Burst Effects at High Overpressure*. Los Alamos, New Mexico: Defense Atomic Support Agency (AEC)(DASA 2506).

Brown, R. A., et al. 1974. *SMOBSMOD (Strategic Mobility Simulation Model): User's Manual*. Bethesda, Maryland: U.S. Army Concepts Analysis Agency (LD-31060).

Brownlee, K. A. 1948. *Industrial Experimentation*. New York: Chemical Publishing Company, Inc.

Brownlee, K. A. 1960. *Statistical Theory and Methodology in Science and Engineering*. New York: John Wiley and Sons, Inc.

Burington, R., and May, D. 1958. *Handbook of Probability and Statistics With Tables*. Sandusky, Ohio: Handbook Publishers, Inc.

Catalog of War Gaming and Military Simulation Models. 1977. 7th ed. Washington, D.C.: Joint Chiefs of Staff, Studies, Analysis and Gaming Agency (SAGA-180-77).

Christ, C. F. 1966. *Econometric Models and Methods*. New York: John Wiley and Sons, Inc.

Close, A. F., and Gillen, C. A. 1969. *Description of the Computer Program for METRIC-A Multi-Echelon Technique for Recoverable Item Control*. Santa Monica, California: Rand Corporation (AD 686056).

Cochran, W. G., and Cox, G. M. 1957. *Experimental Designs*. New York: John Wiley and Sons, Inc.

Cochran, William G. 1953. *Sampling Techniques.* New York: John Wiley and Sons, Inc. (2d. ed. 1963).

Cochrane, J. L., and Zeleny, eds. *Multiple Criteria Decision Making.* Columbia, South Carolina: University of South Carolina Press.

Cole, J. C., and Dorr, A. J. 1974. *The OASIS-II Program. Volume V: OASIS-II Listings.* La Jolla, California: Science Applications, Inc. (SAI-74-251-LJ).

Cole, J. C., and Dorr, A. J. 1974. *CARMONETTE: General Description.* McLean, Virginia: General Research Corporation (AD A007 843).

Combat II Model: A Step Toward an Integrated Combat Assessment, The. 1974. Vienna, Virginia: Braddock, Dunn, and McDonald (LD-36936MA).

Conover, W. J. 1971. *Practical Non-Parametric Statistics.* New York: John Wiley and Sons, Inc.

Cramer, H. 1946. *Mathematical Methods of Statistics.* Princeton, New Jersey: Princeton University Press.

Curtis, W., and Smith, K. L. 1968. *Preliminary Test Design Procedures.* Albuquerque, New Mexico: Sandia Corporation.

Davies, O. L., ed. 1956. *The Design and Analysis of Industrial Experiments.* New York: Hafner Publishing Company.

Davis, Harold. 1974. *Maintenance of Strategic Models for Arms Control Analysis - Description of Strategic International Relations Nuclear Exchange Model.* Chicago, Illinois: Academy for Interscience Methodology (AD-780-490).

Department of Defense. 1973. *Life Cycle Costing Guide for System Acquisition (Interim).* Washington, D.C.: U.S. Government Printing Office (008-000-00184-7).

DeSobrino, R. 1969. *A Comparison of Two Strategic Forces Evaluation Models.* Menlo Park, California: Stanford Research Institute (SED-RM-5205-53).

DIVWAG Model Documentation. 1976. Fort Leavenworth, Kansas: Army Combat Development Activity (LD 40635MA).

Dixon, W. T., and Massey, F. T. 1957. *Introduction to Statistical Analysis.* New York: McGraw-Hill Book Company.

Dixon, W. T., and Massey, F. T. 1957. *An Introduction to Statistical Analysis* (2d ed.). New York: McGraw-Hill Book Company.

Dockery, J., Leiser, M., and Aitben, M. 1976. A Theater-Level Simulation With COMO-III. In *Computers and Operations Research,* Volume 3, pp. 57-71. New York: Pergamon Press.

Dolins, L. P. 1970. *The IDA Level Defense Economic Model: An Interim Summary.* Arlington, Virginia: Institute for Defense Analysis (AD 711 553).

Donelson, J., et al. 1976. *IDA Strategic Nuclear Exchange Model (IDASNEM).* Washington, D.C.: Institute for Defense Analysis (AD B020 653L).

Drake. M. K., and Frike, M. P., eds. 1974. *Nuclear Weapon Effects Handbook for Long Range Research and Development Program.* La Jolla, California: Science Applications, Inc.

Draper, N., and Smith, H. 1981. *Applied Regression Analysis* (2d ed.). New York: John Wiley and Sons, Inc.

Durbin, James, and Watson, G. S. 1950. Testing for Serial Correlation in Least Squares Regression. I. *Biometrika* 37 (December), pp. 409-428.

Durbin, James, and Watson, G. S. 1951. Testing for Serial Correlation in Least Squares Regression. II. *Biometrika* 38 (June), pp. 159-178.

Durbin, James. 1960. Estimation of Parameters in Time-series Regression Models, *Journal of the Royal Statistical Society,* Series B (Methodological), Volume 22, No. 1 (January), pp. 139-153.

Edited by McMaster for the Society for Nondestructive Testing. 1963. *Nondestructive Testing Handbook.* New York: The Ronald Press Company.

Emerson, D. E. 1969. *UNCLE-A New Force-Exchange Model for Analyzing Strategic Uncertainty Levels.* Santa Monica, California: Rand Corporation (R-430-PR).

Enthoven, A. C., and Smith, K. W. 1971. *How Much Is Enough?* New York: Harper and Row.

Everett, Hugh M. 1973. Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources. *Operations Research* 11:339-425.

Fain, J. 1975. The Lanchester Equations and Historical Warfare: An Analysis of Sixty World War II Land Engagements. In *Proceedings of the 34th Military Operations Research Symposium.* Alexandria, Virginia: Military Operations Research Society (AD C002 913).

Farrell, R. L., Cherry, W. P., and Ahrens, R. A. 1976. *A Preliminary Design and Resource Estimate for a Combined-Arms Simulation Model (CASM). Volume I: Summary Report.* Ann Arbor, Michigan: Vector Research (AD B010 532L).

Fish, J. R. 1975. *ATACM:ACDA Tactical Air Campaign Model.* Arlington, Virginia: Ketron (AD B008 645).

Fisher, R. A. 1958. *Statistical Methods for Research Workers.* New York: Hafner Publishing Company.

Fisher, R. A. 1960. *The Design of Experiments.* New York: Hafner Publishing Company.

Fisher and Yates. 1963. *Statistical Tables for Biological, Agricultural, and Medical Research.* New York: Hafner Publishing Company.

Forsythe, G. E. 1957. Generation and Use of Orthogonal Polynomials for Data Fitting With a Digital Computer, *Journal of the Society for Industrial and Applied Mathematics* 5:74-88.

FOURCE, Division Level Combat Model. Working papers from *TOS Cost and Operational Effectiveness Analysis.* White Sands Missile Range, TRADCO Systems Analysis Activity (ATAA-TCA).

Fry, F. C. 1948. *Probability and Its Engineering Uses.* Princeton, New Jersey: Van Nostrand.

Function FEAR: A Set of Subroutines for the UNICORN Model for Rapid Calculation of Lethal Offsets From Nuclear Weapons (U). 1975. La Jolla, California: Science Applications, Inc. (SAI-75-125-LJ).

Green, D. W., et al. 1977. *The SRI-WEFA Soviet Econometric Model: Phase Three Documentation.* Volume I. Arlington, Virginia: SRI International (AD A043 287).

Groover, P. L., et al. 1978. *Assumptions and Rationale Used To Calculate an Arsenals Relative Force Size.* Springfield, Virginia: Science Applications, Inc. (SAI-78-03-POST).

Hannan, E. J. 1957. Testing for Serial Correlation in Least Squares Regression, *Biometrika* 44, Parts 1 and 2 (June), pp. 57-66.

Hannan, E. J. 1965. The Estimation of Relationships Involving Distributed Lags, *Econometrica*, Volume 33, No. 1 (January), pp. 206-224.

Hartley, D. S., and Cockrell, A. A. 1975. *An Examination of a Distribution of TAC Contender Solutions.* Washington, D.C.: National Military Command System Support Center (AD A014 426).

Hausrath, A. H. 1971. *Venture Simulation in War, Business, and Politics.* New York: McGraw-Hill Book Company.

Hawkins, J. H. 1976. *The AMSAR War Game (AMSWAG) Computer Combat Simulation.* Aberdeen Proving Ground, Maryland: Army Material Systems Analysis Activity (AD A028 052).

Hengle, J. E. 1976. Quantitative Methodology for Evaluating Strategic Command Control and Communications Systems. *Proceedings of the 37th Military Operations Research Symposium (MORS).* Alexandria, Virginia: Military Operations Research Society (AD C009 755L).

Hess, R. W. 1974. *BOMTAN: A Model for Estimating the Annual Cost of Bomber and Tanker Squadrons.* Santa Monica, California: Rand Corporation (LD 31844MA).

Hicks, C. R. 1964. *Fundamental Concepts in the Design of Experiments.* New York: Hold, Rinehart, and Winston.

Hierarchy of Combat Analysis Models, A. 1973. McLean, Virginia: General Research Corporation (AD 906 982).

Hillerman, Hall H. 1972. *The Theoretical Basis of the CODE 50 Nuclear Exchange Model.* Arlington, Virginia: Center for Naval Analysis (AD A043 377).

Hitch, C. J., and McKean, R. N. 1978. *The Economics of Defense in the Nuclear Age.* Chapter 9. New York: Atheneum.

Hoel, P. G. 1962. *Introduction to Mathematical Statistics.* New York: John Wiley and Sons, Inc.

Honig, J., et al. 1971. *Review of Selected Army Models.* Washington, D.C.: Assistant Vice Chief of Staff (AD 887 175L).

Hood, J. W., and Terry, W. E. 1974. *Strategic Arms Interactions: 1945-1961.* Monterey, California: U.S. Naval Post- graduate School (AD A001 264).

Hooper, John W., and Zellner, Arnold. 1961. The Error of Forecast for Multivariate Regression Models, *Econometrica,* Volume 29, No. 4 (October), pp. 544-555.

Hopgood, Roy E. 1965. *Military Planning Game.* Monterey, California: U.S. Naval Postgraduate School (AD 475 372).

Huff, Darrell. 1954. *How To Lie With Statistics.* New York: Norton.

Hugus, D. K. 1976. *A Review of the Compilation of the DYNTACS(x) Data Base for the*

HELLFIRE COEA. Fort Leavenworth, Kansas: Army Combat Development Activity (AD A034 919).

Ignizio, J. 1977. *Goal Programming and Extensions.* New York: Lexington.

Integrated Nuclear and Conventional Theater Warfare Simulation (INWARS). Level III Specifications. Volume I: Introduction. McLean, Virginia: BDM Corporation (BDM/W-78-402-TR).

Johnson, N. L., and Leone, F. C. 1964. *Statistical and Experimental Design.* New York: John Wiley and Sons, Inc.

Johnson, P. O. 1950. *Statistical Methods in Research.* New York: Prentice-Hall, Inc.

Judnick, Wm. E. 1978. Use of Experimental Design With Theater-Level Models (U). In *Proceedings of the 39th Military Operations Research Symposium (June 1977).* Alexandria, Virginia: Military Operations Research Society.

Kahan, J. H. 1975. *Security in the Nuclear Age-Developing U.S. Strategic Arms Policy.* Washington, D.C.: The Brookings Institute.

Karr, A. F. 1973. *Adaptive and Behavioral Strategies in Two-Person, Two-Move, Two-Action, Zero-Sum Games.* Arlington, Virginia: Institute for Defense Analysis (IDA Paper P-993).

Karr, A. F. 1976a. *On the CONAF Evaluation Model.* Arlington, Virginia: Institute for Defense Analysis (LD 37409MA).

Karr, A. F. 1977. *Review and Critique of the VECTOR-II Combat Model.* Arlington, Virginia: Institute for Defense Analysis (IDA Paper P-1315).

Kelleher, G. J. 1965. *A Damage Limiting Shelter-Allocation Strategy.* Arlington, Virginia: Institute for Defense Analysis (AD 615779).

Kempthorne, O. 1960. *The Design and Analysis of Experiments.* New York: John Wiley and Sons, Inc. (second printing).

Kempthorne, O. 1962. *Design and Analysis of Experiments.* New York: John Wiley and Sons, Inc.

Kennedy, W. R., et al. 1977. *DIVLEV War Game Model Computer Program.* Aberdeen Proving Ground, Maryland: U.S. Army Material Systems Analysis Activity.

Keynes, John Maynard. 1936. *The General Theory of Employment, Interest, and Money.* London: MacMillan.

Koopman, Bernard O. 1970. A Study of the Logical Basis of Combat Simulation. *Operations Research 18:855-882.*

Koopmans, Tjalling C., ed. 1950. *Statistical Inference in Dynamic Economic Models.* Cowles Commission Monograph 10. New York: John Wiley and Sons, Inc.

Lanchester, F. W. 1914. Aircraft in Warfare: The Dawn of the Fourth Arm, No. 5: The Principle of Concentration. Reprinted 1956 in *The World of Mathematics,* Volume 4, ed. Newman, J., New York: Simon and Schuster. pp. 2138-2148.

Larson, Harold J., and Bancroft. T. A. 1963a. Biases in Prediction by Regression for Certain Incompletely Specified Models, *Biometrika,* Volume 50, Parts 3 and 4 (December), pp. 391-402.

Leake, L. A., Tiede, R. V., Whipple, S., Jr. 1970. *Information Flow and Combat Effectiveness. Part II: Methodology for Evaluating the Effectiveness and Cost of the Tactical Operations System (TOS).* McLean, Virginia: Research Analysis Corporation (RAC-R-100).

Livingston, I., and Groover, P. 1974. *The HALL Model for the Analysis of Aircraft Survivability Prior to Penetration.* Denver, Colorado: Science Applications, Inc. (SAI 74-001-DEN).

Low, L. J., and Means, E. H. 1976. *BALFRAM (Balanced Force Requirements Analysis Methodology. Volume I: Seminar Guide.* Menlo Park, California: Stanford Research Institute (AD B013 416L).

Mailart, J. D., et al. 1974. *Navy Modified Logistics Composite Model (LCOM) User's Manual.* Los Angeles, California: CACI, Incorporated-Federal (AD A006 975).

Massey, H. G. 1977. *Introduction to the USAF Total Force Cost Model.* Santa Monica, California: Rand Corporation (AD A042 460).

Mathematical Formulations of the Arsenal Exchange Model: Revision 7. 1975. Denver, Colorado: Martin Marietta Corporation.

MEECN System Simulator Capabilities Description. 1974. Washington, D.C.: Defense Communications Agency (TP 960-TP-74-40).

Mendenhall, W. 1968. *Introduction to Linear Models and the Design and Analysis of Experiments.* Belmont, California:
Wadsworth Publishing Company, Inc.

Mobile Missile Survival Against a Ballistic Salvo - The Effects of Position Uncertainty and Other Parameters. 1958. Santa Monica, California: Rand Corporation (RM-2743).

Models of the U.S. Army Worldwide Logistics System. 1977. Volumes 2A and 3. Vienna, Virginia: BDM Corporation (LD 3829OMC and AD-A040-804).

Moroney, M. J. 1956. *Facts From Figures.* Baltimore, Maryland: Penguin Books.

Neu, R. M. 1974. *Attacking Hardened Air Bases (AHAB): A Decision Analysis Aid for the Tactical Commander.* Santa Monica, California: Rand Corporation (AD 786 879).

NMCSSC Strategic War Gaming Support With the Quick-Reacting War Gaming System (QUICK). 1974. Arlington, Virginia: Defense Communications Agency (TM 90-74).

Novick, D. 1974. *Program Budgeting in the Department of Defense.* Santa Monica, California: Rand Corporation (605 388).

Operation Test and Evaluation Simulation Guide. 1974. Vienna, Virginia: Braddock, Dunn, and McDonald. *Volume I: Summary* (AD A001 692). *Volume II: Hardware* (AD A001 693). *Volume III: Classified Simulator Devices (U)* (AD C000 021). *Volume IV: Software Simulators* (AD A001 694).

Ostle, B. 1963. *Statistics in Research.* Ames, Iowa: Iowa State University Press.

Parzan, E. 1960. *Modern Probability Theory and Its Application.* New York: John Wiley and Sons, Inc.

Paulson, R. M., et al. 1971. *Using Logistics Models in System Design and Early Support Planning.* Santa Monica, California: Rand Corporation (AD 724 690).

Pearson, E. S., and Hartley, H. O. 1951. *Biometrika,* Vol. 39, pp. 115-122.

Phelps, M. W., et al. 1976. *Generalized Decomposition Procedures for Strategic Balance Calculations.* Denver, Colorado: Science Applications, Inc. (SAI-76-084-DEN).

Pietrewicz, P. A., et al. 1978. *Development of a Methodology for Solving the Combat Potential Delivery Problem: A Final Research Report.* Denver, Colorado: Science Applications, Inc. (SAI-78-115-DEN).

Preliminary Design and Resource Estimate for Combined Arms Simulation Model (CASM). Volume I: Executive Summary. 1976. Vienna, Virginia: BDM Corporation (AD B010 526L).

Quade, E. S. 1970. *Analysis for Military Decision.* Chapter 5. New York: American Elsevier Publishing Co.

Raiffa, Howard, and Schlaifer, Robert. 1961. *Applied Statistical Decision Theory.* Boston, Division of Research, Graduate School of Business Administration, Harvard University.

Robinson, R. C., McDowell, A. C., and Rowan, J. W. 1971. *Simulation and Gaming Methods for Analysis of Logistics, Part II (SIMALOG II): Capability Analysis System.* McLean, Virginia: Research Analysis Corporation (LD 16779R).

Savage, Leonard J. 1954. *The Foundations of Statistics.* New York: John Wiley and Sons, Inc.

Savage, Leonard J. 1962. *The Foundations of Statistical Inference.* New York: Wiley, and London, Methuen.

Sheffe, H. 1959. *The Analysis of Variance.* New York: John Wiley and Sons, Inc.

Shubik, Martin. 1975a. *Games for Society, Business, and War: Towards a Theory of Gaming.* New York: Elsevier Scientific Publishing Company.

Shubik, Martin. 1975b. *The Uses and Methods of Gaming.* New York: Elsevier Scientific Publishing Company.

Siegel, S. 1956. *Nonparametric Statistics for the Behavioral Sciences.* New York: McGraw-Hill Book Company.

Simon, Herbert A. 1955. Prediction and Hindsight as Confirmatory Evidence, *Philosophy of Science,* Volume 22, No. 3 (July), pp. 227-230.

Slonim, M. J. 1960. *Sampling in a Nutshell.* New York: Simon and Schuster.

Spaulding, S. L. 1971. *User's Manual for Bonder/IUA Combat Model.* Washington, D.C.: Assistant Vice Chief of Staff, Weapons Systems Analysis Directorate (LD 32756MA).

Spaulding, S. L., Samuel, A. H., Judnick, W. E., et al. 1977. Tactical Force Mix Versatility Planning with VECTOR-I and UNICORN (U). In *Proceedings of the 38th Military Operations Research Symposium (June 1976).* Alexandria, Virginia: Military Operations Research Society (AD C012 958L).

Staff Officer's Field Manual Nuclear Weapons Employment Doctrine and Procedures. 1968. Washington, D.C.: Department of the Army and Navy.

Stockfisch, J. A. 1975. *Models, Data, and War: A Critique of the Study of Conventional Forces.* Santa Monica, California: Rand Corporation (R-1526-PR).

Tactical Warfare Simulation Evaluation and Analysis Systems (TWSEAS). 1976. Quantico, Virginia: Marine Corps Development and Education Command (AD B007 867L).

Tang, P. C. 1938. "The Power Function of the Analysis of Variance." Tests With Tables and Illustrations of Their Use. *St atistical Research* Memoirs, Volume 2, pp. 126-149.

Tank, Anti-Tank, and Assault Weapon Requirements Study (U). 1968. Fort Knox, Kentucky: U.S. Army Combat Development Command, Armor Agency (ACN 7356).

Tate, Merle W., and Clelland, Richard C. 1957. *Nonparametric and Shortcut Statistics.* Danville, Illinois: Interstate Printers and Publishers, Inc.

Taylor, J. G. 1975. A Tutorial on Lanchester-Type Models of Warfare. In *Proceedings of the 35th Military Operations Research Symposium (July 1975).* Alexandria, Virginia: Military Operations Research Society (AD C005 592).

Techniques for Evaluating Strategic Bomber Penetration: An Aggregated Penetration Model--PENEX. 1972. Washington, D.C.: United States Air Force Assistant Chief of Staff.

TEMPER, Volume II: The Theory of the Model. 1965. Bedford, Massachusetts: Raytheon (AD 471 458).

Thomas, Jerry. 1975. *Tartarus Model Sensitivity Experiment.* Bethseda, Maryland: Army Concepts Analysis Agency (AD A018 896).

Thrall, R. M., Tsokos, C. P., Turner, J. C., eds. 1976. *Proceedings of the Workshop on Decision Information for Tactical Command and Control.* Houston, Texas: Robert M. Thrall (AD A052 439).

Tiede, Roland V., and Leake, Lewis A. 1971. A Method for Evaluating the Combat Effectiveness of a Tactical Information System in a Field Army. *Operations Research.* 19:587-604.

Tinall, John B., et al. 1973. *A Primer in Strategic Command and Control Communications* Colorado Springs, Colorado: Air Force Academy (AD 786-034).

VECTOR-II System for Simulation of Theater-Level Combat. Part I: User's Manual. 1978. Washington, D.C.: Joint Chiefs of Staff (Computer System Manual CSM UM 244-78).

Villars, D. S. 1951. *Statistical Design and Analysis of Experiments for Development Research.* Dubuque, Iowa: Wm. C. Brown.

Wald, Abraham. 1950. *Statistical Decision Functions.* New York: John Wiley and Sons, Inc.

Walker, G. A. 1974. *Life Cycle/System Effectiveness Evaluation and Criteria.* Seattle, Washington: Boeing Company (AD 916 001).

Walker, L. K., et al. 1976. *Comparison and Evaluation of Four Theater-Level Models: CEM IV, IDAGAM I, LULEJIAN I, VECTOR I.* Arlington, Virginia: Weapons Systems Evaluation Group (AD A036 325).

Walther, Harry J. 1975. *Catalog of War Gaming and Military Simulation Models.* Washington, D.C.: Organization of the Joint Chiefs of Staff (SAGA-236-75).

Watson, G. S., and Hannan, E. J. 1956. Serial Correlation in Regression Analysis. II, *Biometrika* 43, Parts 3 and 4 (December), pp. 436-448.

Weaver, Warren. 1963. *Lady Luck.* Garden City, New York: Doubleday.

Wetzler, Elliot. *The Structure of the IDA Civil Defense Model.* Arlington, Virginia: Institute for Defense Analysis (AD 717 098).

White, W., et al. 1975. *DIVOPS. A Division-Level Combined Arms Engagement Model. Version 3. Volume III: Planner/User Manual.* Fort Leavenworth, Kansas: BDM Services Company (LD 36265MD).

Willard, D. 1962. *Lanchester as Force in History; An Analysis of Land Battles of the Years 1618-1905.* McLean, Virginia: Research Analysis Corporation (RAC-TP-74).

Wilson, Andrew. 1970. *The Bomb and the Computer.* New York: Dell Publishing Company.

Wilson, E. Bright, Jr. 1952. *An Introduction to Scientific Research.* New York: McGraw-Hill Book Company.

Winer, B. J. 1962. *Statistical Principles in Experimental Design.* New York: McGraw-Hill Book Company.

Yergin, Daniel. 1977. *Shattered Peace.* Boston, Massachusetts: Houghton Mifflin Company.

Young, H. D. 1962. *Statistical Treatment of Experimental Data.* New York: McGraw-Hill Book Company.

Zellner, Arnold, and Tiao, George C. 1964. Bayesian Analysis of the Regression Model With Autocorrelated Errors, *Journal of the American Statistical Association,* Vol. 59, No. 307 (September), pp. 763-778.

Zimmerman, R. E. 1956. A Monte Carlo Model for Military Analyses. In *Operation Research for Management.* Volume 2, eds. McCloskey and Coppinger. Baltimore, Maryland: Johns Hopkins.